HELSINGIN YLIOPISTO
Pro gradu -tutkielma

Niko Kaitarinne

Singulaariarvohajotelma ja pseudoinverssi

Matematiikan ja tilastotieteen laitos
Matematiikka
Helmikuu 2012

Helsingin yliopisto

Matematiikan ja tilastotieteen laitos

KAITARINNE, NIKO: Singulaariarvohajotelma ja pseudoinverssi

Pro gradu -tutkielma, 45 s.

Matematiikka

Helmikuu 2012

Tiivistelmä

Tässä tutkielmassa tarkastellaan singulaariarvohajotelmaa ja sen sovelluksena pseudoinverssiä. Lukija johdatellaan aiheeseen esittelemällä ensin joitakin tarvittavia esitietoja. Tämän jälkeen esittelemme unitaariset matriisit, jotka ovat yleistys ortogonaalisista matriiseista. Unitaaristen matriisien avulla esittelemme myös unitaarisen similaarisuuden käsitteen sekä Schurin hajotelmaksi kutsutun lauseen. Seuraavaksi esittelemme normaalit matriisit ja käsittelemme niiden tärkeimmät ominaisuudet. Vielä ennen tutkielman aiheeseen pääsyä käsittelemme hermiittiset matriisit, jotka ovat symmetristen matriisien yleistys, ja definiitit matriisit, jotka ovat hermiittisten matriisien eräänlainen osajoukko. Definiittien matriisien avulla käsitellään singulaariarvot sekä niiden avulla saatava singulaariarvohajotelma. Lopulta määritellään pseudoinverssi ja käydään läpi sen ominaisuudet, jonka jälkeen sovelletaan sitä pienimmän neliösumman probleeman ratkaisuun.

Sisältö

1 Johdanto **4**

2 Esitiedot **5**
2.1 Perusmääritelmiä 5
2.2 Ominaisarvot ja ominaisvektorit 6
2.3 Similaarisuus 7
2.4 Sisätulo ja ortogonaalisuus 7
2.5 Lineaarisista yhtälöryhmistä 9

3 Unitaariset ja normaalit matriisit **11**
3.1 Unitaariset matriisit 11
3.2 Normaalit matriisit 16

4 Hermiittiset ja positiivisesti (semi)definiitit matriisit **20**
4.1 Hermiittiset matriisit 20
4.2 Positiivisesti (semi)definiitit matriisit 23

5 Singulaariarvohajotelma **27**
5.1 Singulaariarvot 27
5.2 Singulaariarvohajotelma 29

6 Pseudoinverssi ja pienimmän neliösumman probleema **33**
6.1 Pseudoinverssi 33
6.2 Pienimmän neliösumman probleema 41

Viitteet **46**

1 Johdanto

Tässä Pro Gradu –tutkielmassa tutkielmassa tarkastellaan singulaariarvohajotelmaa, joka on hajotelma mille tahansa matriisille. Vaikka kyseessä ei ole yksikäsitteinen hajotelma, on se kuitenkin tärkeä osa lineaarialgebran teoriaa sekä monien sovellusten lähtökohta. Lähtötietoina lukijalta oletetaan lineaarialgebran alkeiden osaamista, ja joitakin niistä käydään läpi myös toisessa luvussa. Lisäksi lukijan tulee osata kompleksilukujen alkeet.

Luvussa 2 käydään läpi tutkielmassa tarvittavia esitietoja. Joitakin lineaarialgebran perusteita kuten matriisin ydin ja kuva sekä ominaisarvot on esitelty lukijalle lähinnä muistuksena ensimmäisiltä lineaarialgebran kursseilta. Jotkin aihealueet voivat olla lukijoille ennestään tuntemattomia, mutta niiden omaksumisen ei pitäisi aiheuttaa ongelmia. Tämän luvun lähteenä on käytetty kirjaa [3].

Luvussa 3 käsitellään unitaarisia ja normaaleja matriiseja. Ensimmäisessä aliluvussa esitellään unitaariset matriisit sekä niiden ominaisuudet. Lisäksi käsitellään unitaarisen similaarisuuden käsite sekä Schurin hajotelma. Kyseinen aliluvun määritelmät ja lauseet perustuvat kirjan [1] luvuissa 2.1-2.3 esitettyihin määritelmiin ja lauseisiin lukuunottamatta lauseen 3.1 kohtaa (7), joka perustuu kirjan [3, s. 189] lauseeseen 5.6.3. Toisessa aliluvussa käydään läpi normaalit matriisit ja niiden ominaisuudet. Tämä aliluku perustuu suurimmilta osin kirjan [3] lukuihin 5.2 ja 5.6.

Luvussa 4 tarkastellaan hermiittisiä matriiseja sekä niiden avulla johdettuja definiittejä matriiseja. Samalla siirrytään kohti singulaariarvon käsitettä, jonka perusteet löytyvät definiiteistä matriiseista. Hermiittisten matriisien osio perustuu kirjan [1] kappaleesta 4.1 ja kirjan [3] kappaleesta 5.3 löytyviin määritelmiin ja lauseisiin. Definiittien matriisien kappaleen lähteenä on käytetty kirjan [3] lukuja 5.3-5.4.

Luku 5 on tutkielman pääluku. Siinä käydään läpi singulaariarvohajotelman käsite. Ensin esitellään singulaariarvot, jonka jälkeen voimme muodostaa singulaariarvohajotelman. Singulaariarvoja käsittelevä aliluku perustuu kirjan [3] luvun 5.4 loppuosaan. Aliluku singulaariarvohajotelmasta perustuu kirjan [3] lukujen 5.7 ja 12.8 aihetta käsitteleviin osiohin.

Luvussa 6 määrittelemme pseudoinverssin ja joitakin sen ominaisuuksia sekä käytämme sitä pienimmän neliösumman parhaan ratkaisun löytämiseen. Löydämme myös pseudoinverssille eksaktin muodon singulaariarvohajotelmaa hyväksi käyttäen. Tämän luvun lähteenä on käytetty kirjan [3] lukuja 12.8 ja 12.9.

2 Esitiedot

Tässä luvussa käsittelemme luettelonomaisesti tutkielman määritelmissä ja lauseiden todistuksissa tarvittavia apulauseita. Kaikkien lauseiden todistukset sivuutetaan tässä luvussa.

Käsitellään ensin hieman merkintöjä, joita käytämme tässä tutkielmassa. Nollamatriisia merkitään symbolilla $\mathbf{0}$ ja nollavektoreita merkitsemme symbolilla Θ, emmekä merkitse niiden kokoa, jos ei sekaantumisen vaaraa ole. Diagonaalimatriisit lyhennetään merkinnällä $D = \operatorname{diag}(a_1, a_2, \ldots, a_n)$, jossa diagonaalialkiot ovat sulkujen sisällä lueteltuina järjestyksessä vasemmasta yläkulmasta oikeaan alakulmaan.

2.1 Perusmääritelmiä

Ensiksi käsittelemme joitakin konjugaattitranspoosiin liittyviä määritelmiä ja lauseita sekä muistutetaan mieleen, mikä on matriisin ydin ja kuva. Matriisin ytimeen ja kuvaan liittyviä lauseita käsitellään lisää myöhemmin tässä luvussa.

Määritelmä 2.1. Olkoon $A = \begin{bmatrix} a_{ij} \end{bmatrix} \in \mathbb{C}^{m \times n}$. Tällöin matriisia $B = \overline{A^T} = \begin{bmatrix} \overline{a_{ji}} \end{bmatrix}$ kutsutaan *matriisin A konjugaattitranspoosiksi* ja merkitään $B = A^*$.

Esimerkki 2.1. Olkoon

$$A = \begin{bmatrix} 2-i & -i & 1 \\ 3 & 1+i & 2i \end{bmatrix}.$$

Tällöin

$$A^* = \begin{bmatrix} 2+i & 3 \\ i & 1-i \\ 1 & -2i \end{bmatrix}.$$

Lause 2.1. *Konjugaattitranspoosille pätee ehdot*

(1) $(A^*)^* = A$,

(2) $(A+B)^* = A^* + B^*$,

(3) $(aA)^* = \overline{a}A^*$,

(4) $(AB)^* = B^*A^*$,

(5) $(A^{-1})^* = (A^*)^{-1}$, *jos matriisi A on kääntyvä,*

kun A ja B ovat sopivan kokoisia matriiseja ja $a \in \mathbb{C}$.

Määritelmä 2.2. Olkoon $A \in \mathbb{C}^{m \times n}$. Tällöin joukkoa

$$\operatorname{Ker} A = \{\boldsymbol{x} \in \mathbb{C}^n | A\boldsymbol{x} = \Theta\},$$

kutsutaan *matriisin A nolla-avaruudeksi tai ytimeksi* ja joukkoa

$$\operatorname{Ran} A = \{\boldsymbol{y} \in \mathbb{C}^n | \boldsymbol{y} = A\boldsymbol{x} \text{ jollekin } \boldsymbol{x} \in \mathbb{C}^n\},$$

kutsutaan *matriisin A kuva-avaruudeksi tai kuvaksi.*

Matriisin A kuva voidaan määritellä myös lineaarikombinaation avulla. Tällöin $\boldsymbol{y} \in \operatorname{Ran} A$, jos ja vain jos $\boldsymbol{y}$ voidaan esittää matriisin A sarakevektoreiden lineaarikombinaationa. Matriisin kuva-avaruuden dimensiota merkitään joskus myös

$$\dim[\operatorname{Ran} A] = \operatorname{rank} A.$$

Esitämme vielä kaksi lausetta liittyen dimenssioihin.

Lause 2.2. *Olkoon $A \in \mathbb{C}^{m \times n}$. Tällöin*

$$\operatorname{rank} A = \operatorname{rank} A^T = \operatorname{rank} A^*$$

Lause 2.3. *Olkoon V ja W äärellisulotteisia vektoriavaruuksia. Jos $V \subseteq W$ ja $\dim V = \dim W$, niin $V = W$.*

2.2 Ominaisarvot ja ominaisvektorit

Seuraavaksi käsittelemme joitakin ominaisarvoihin ja ominaisvektoreihin liittyviä ominaisuuksia.

Määritelmä 2.3. Olkoon $A \in \mathbb{C}^{n \times n}$. Skalaaria $\lambda \in \mathbb{C}$ kutsutaan *matriisin A ominaisarvoksi*, jos se toteuttaa yhtälön

$$A\boldsymbol{x} = \lambda \boldsymbol{x},$$

kun $\boldsymbol{x} \in \mathbb{C}^n$ ja $\boldsymbol{x} \neq \Theta$. Tällöin vektoria $\boldsymbol{x}$ kutsutaan *ominaisarvoa λ vastaavaksi ominaisvektoriksi.*

On tärkeää huomata, että kompleksilukuarvoisten matriisien tapauksessa ominaisarvoja löytyy aina n kappaletta, kun jokaista ominaisarvoa otetaan mukaan sen algebrallisen kertaluvun osoittama määrä. Tämä helpottaa niiden käsittelyä verrattuna reaalilukuarvoisiin matriiseihin, joilta ei välttämättä löydy yhtään reaalista ominaisarvoa.

Lause 2.4. *Erisuuruisiin ominaisarvoihin liittyvät ominaisvektorit ovat aina lineaarisesti riippumattomia.*

Määritelmä 2.4. Olkoot $A \in \mathbb{C}^{n \times n}$ ja $\lambda \in \mathbb{C}$ sen ominaisarvo. Tällöin joukkoa

$$A_\lambda = \{\boldsymbol{x} \in \mathbb{C}^n | A\boldsymbol{x} = \lambda \boldsymbol{x}\}$$

kutsutaan *ominaisarvoon λ liittyväksi ominaisavaruudeksi.*

2.3 Similaarisuus

Seuraavaksi määrittelemme similaariset matriisit ja käsittelemme kaksi niitä koskevaa lausetta.

Määritelmä 2.5. Matriiseja $A \in \mathbb{C}^{n \times n}$ ja $B \in \mathbb{C}^{n \times n}$ kutsutaan keskenään *similaarisiksi matriiseiksi*, jos

$$A = PBP^{-1},$$

jollakin kääntyvällä matriisilla $P \in \mathbb{C}^{n \times n}$. Jos matriisi B on diagonaalimatriisi, niin kutsutaan matriisia A *diagonalisoituvaksi.*

Lause 2.5. *Matriisit $A \in \mathbb{C}^{n \times n}$ ja $B \in \mathbb{C}^{n \times n}$ ovat similaarisia, jos ja vain jos niillä on samat ominaisarvot.*

Lause 2.6. *Olkoon $A \in \mathbb{C}^{n \times n}$ ja olkoot $\lambda_1, \lambda_2, \ldots, \lambda_n$ sen ominaisarvot. Tällöin matriisi A on diagonalisoituva, jos ja vain jos on olemassa matriisin A ominaisarvoja vastaavat lineaarisesti riippumattomat ominaisvektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_n}$, jotka muodostavat vektoriavaruuden $\mathbb{C}^n$ kannan. Lisäksi tällöin $D = diag(\lambda_i)$, kun $i = 1, 2, \ldots, n$, ja*

$$P = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \cdots & \boldsymbol{x_n} \end{bmatrix}.$$

2.4 Sisätulo ja ortogonaalisuus

Tässä luvussa käsittelemme ensin yleisellä tasolla sisätuloa ja normia. Tämän jälkeen siirrymme käsittelemään ortogonalisuutta sekä ortonormaaleja kantoja.

Määrittelemme sisätulon vain vektoriavaruudelle $\mathbb{C}^n$, mutta se voitaisiin määritellä myös yleisesti vektoriavaruudelle V. Tämä ei kuitenkaan ole tarpeellista, sillä sisätulot on mielekästä määritellä vain kompleksi– tai reaalilukuarvoisille vektoriavaruuksille.

Määritelmä 2.6. Tarkastellaan vektoriavaruutta $\mathbb{C}^n$, jonka kerroinkuntana on $\mathbb{C}$. Tällöin kuvausta $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ sanotaan vektoriavaruuden $\mathbb{C}^n$ *sisätuloksi*, jos ehdot

(1) $\langle a\boldsymbol{x} + b\boldsymbol{y}, \boldsymbol{z} \rangle = a \langle \boldsymbol{x}, \boldsymbol{z} \rangle + b \langle \boldsymbol{y}, \boldsymbol{z} \rangle$,

(2) $\overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle} = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$,

(3) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \in \mathbb{R}$ ja $\langle \boldsymbol{x}, \boldsymbol{x} \rangle > 0$, kun $\boldsymbol{x} \neq \Theta$,

ovat voimassa kaikille $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{C}^n$ ja $a, b \in \mathbb{C}$. Vektoriavaruutta, jossa on määritelty sisätulo, kutsutaan *sisätuloavaruudeksi.*

Lause 2.7. *Olkoon $\mathbb{C}^n$ sisätuloavaruus. Tällöin*

(4) $\langle \boldsymbol{x}, a\boldsymbol{y} + b\boldsymbol{z} \rangle = \bar{a} \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \bar{b} \langle \boldsymbol{x}, \boldsymbol{z} \rangle$,

(5) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, *jos ja vain jos* $\boldsymbol{x} = \Theta$ *tai* $\boldsymbol{y} = \Theta$

ovat voimassa kaikille $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ *ja* $a, b \in \mathbb{C}$.

Lause 2.8. *Kuvaus* $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$, *missä*

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^* \boldsymbol{x},$$

on sisätulo.

Tätä kyseistä sisätuloa kutsutaan vektoriavaruuden $\mathbb{C}^n$ *standardisisätuloksi*, ja myöhemmin käsiteltävät sisätulot ovat standardisisätuloja, jollei muuta mainita.

Lause 2.9. *Standardisisätulolle pätee*

$$\langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, A^* \boldsymbol{y} \rangle,$$

kun $A \in \mathbb{C}^{n \times n}$ *ja* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$.

Määritelmä 2.7. Olkoon $\mathbb{C}^n$ sisätuloavaruus ja olkoon $\boldsymbol{x} \in \mathbb{C}^n$. Tällöin vektorin *(euklidinen) normi* saadaan kaavasta

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle},$$

jossa $\langle \cdot, \cdot \rangle$ on vektoriavaruuden $\mathbb{C}^n$ sisätulo.

Määritelmä 2.8. Vektoreita $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ kutsutaan keskenään *ortogonaaliksiksi*, jos $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. Jos lisäksi $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$, niin kutsutaan kyseisiä vektoreita keskenään *ortonormaaleiksi*.

Lisäksi joukkoa S kutsutaan *ortogonaaliseksi (ortonormaaliksi) joukoksi*, jos kaikki sen sisältämät vektorit ovat keskenään ortogonaalisia (ortonormaaleja).

Määritelmä 2.9. Vektoriavaruuden kantaa kutsutaan *ortogonaaliseksi (ortonormaaliksi) kannaksi*, jos sen virittäjävektorit ovat keskenään ortogonaalisia (ortonormaaleja).

Lause 2.10 (Pythagoraan lause)**.** *Olkoon* $\mathbb{C}^n$ *sisätuloavaruus ja olkoot vektorit* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ *ortogonaalisia. Tällöin*

$$\|\boldsymbol{x} \pm \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2.$$

Määritelmä 2.10. Olkoon S sisätuloavaruuden $\mathbb{C}^n$ epätyhjä osajoukko. Tällöin *joukon* S *ortogonaalikomplementiksi* kutsutaan joukkoa

$$S^{\perp} = \{\boldsymbol{x} \in \mathbb{C}^n | \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0, \text{kaikilla } \boldsymbol{y} \in S\}.$$

Lause 2.11. *Jokainen äärellisulotteisen vektoriavaruuden lineaarisesti riippumaton osajoukko voidaan täydentää vektoriavaruuden kannaksi.*

Lause 2.12 (Gram-Schmidt -ortogonalisointialgoritmi)**.** *Muodostakoon vektorit $\boldsymbol{y_1}, \boldsymbol{y_2} \dots \boldsymbol{y_k}$ vektoriavaruuden $\mathbb{C}^n$ lineaarisesti riippumattoman osajoukon. Tällöin saamme muodostettua vektoriavaruuden $\mathbb{C}^n$ ortogonaalisen osajoukon $S = \{\boldsymbol{x_1}, \boldsymbol{x_2} \dots \boldsymbol{x_k}\}$ siten, että $\boldsymbol{x_1} = \boldsymbol{y_1}$ ja*

$$\boldsymbol{x_m} = \boldsymbol{y_m} - \sum_{j=1}^{m-1} \frac{\langle \boldsymbol{y_m}, \boldsymbol{x_j} \rangle}{\langle \boldsymbol{x_j}, \boldsymbol{x_j} \rangle} \boldsymbol{x_j},$$

kun $m = 2, \dots, k$.

Saamme muodostettua vektoreista $\boldsymbol{x_1}, \boldsymbol{x_2} \dots, \boldsymbol{x_k}$ ortonormaalit vektorit $\boldsymbol{z_1}, \boldsymbol{z_2} \dots, \boldsymbol{z_k}$ jakamalla jokaisen vektorin $\boldsymbol{x_i}$ sen omalla normillaan, eli

$$\boldsymbol{z_i} = \frac{\boldsymbol{x_i}}{\|\boldsymbol{x_i}\|},$$

kun $i = 1, 2, \dots, k$.

Lause 2.13. *Jokaisella äärellisulotteisella sisätuloavaruudella on ortonormaali kanta.*

Lause 2.14. *Muodostakoon vektorit $\boldsymbol{x_1}, \boldsymbol{x_2} \dots \boldsymbol{x_n}$ sisätuloavaruuden $\mathbb{C}^n$ ortonormaalin kannan. Tällöin vektori $\boldsymbol{y} \in \mathbb{C}^n$ voidaan esittää yksikäsitteisessä muodossa*

$$\boldsymbol{y} = \sum_{i=1}^{n} \beta_i \boldsymbol{x_i},$$

jossa $\beta_i = \langle \boldsymbol{y}, \boldsymbol{x_i} \rangle$, kun $i = 1, 2, \dots, n$.

Lause 2.15. *Olkoon $A \in \mathbb{C}^{m \times n}$. Tällöin*

$$\operatorname{Ker} A^* = (\operatorname{Ran} A)^{\perp} \qquad \textit{ja} \qquad \operatorname{Ran} A^* = (\operatorname{Ker} A)^{\perp}.$$

2.5 Lineaarisista yhtälöryhmistä

Tässä luvussa kertaamme lineaarisen yhtälöryhmän ja siitä muodostetun matriisiyhtälön peruskäsitteitä kappaletta 6.2 varten.

Määritelmä 2.11. Tarkastellaan lineaarista yhtälöryhmää

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m. \end{cases}$$

Tämä yhtälöryhmä voidaan kirjoittaa matriisien avulla ekvivalenttiin muotoon $A\boldsymbol{x} = \boldsymbol{b}$, jossa

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \text{ja} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Tällöin yhtälöä $A\boldsymbol{x} = \boldsymbol{b}$ kutsutaan *lineaarisen yhtälöryhmän matriisimuodoksi* tai lyhyemmin *lineaariseksi systeemiksi*, ja matriisia A kutsutaan *kerroinmatriisiksi*. Jos lineaarisella systeemillä on sen toteuttava vektori $\boldsymbol{x_1}$, niin tätä vektoria kutsutaan *lineaarisen systeemin ratkaisuvektoriksi*.

Lause 2.16. *Olkoon $A\boldsymbol{x} = \boldsymbol{b}$ lineaarinen systeemi, jossa $A \in \mathbb{C}^{m\times n}$, $\boldsymbol{b} \in \mathbb{C}^m$ ja $\boldsymbol{x} \in \mathbb{C}^n$. Tällöin systeemillä on ratkaisu, jos ja vain jos $\boldsymbol{b} \in \operatorname{Ran} A$.*

Lause 2.17. *Olkoon $A\boldsymbol{x} = \boldsymbol{b}$ lineaarinen systeemi, jossa $A \in \mathbb{C}^{m\times n}$, $\boldsymbol{b} \in \mathbb{C}^m$ ja $\boldsymbol{x} \in \mathbb{C}^n$. Tällöin vektori $\boldsymbol{x}$ on sen ratkaisuvektori, jos ja vain jos*

$$\|A\boldsymbol{x} - \boldsymbol{b}\| = 0.$$

3 Unitaariset ja normaalit matriisit

Tässä luvussa käsittelemme unitaarisia ja normaaleja matriiseja. Ensimmäisessä aliluvussa käsittelemme unitaarisia matriiseja ja niiden ominaisuuksia, jotka ovat suuressa roolissa myöhemmin aliluvussa 5 esitettävässä singulaariarvohajotelmassa. Toisessa aliluvussa käsittelemme normaaleja matriiseja, jotka ovat johdattelua aliluvussa 4 käsiteltäviin hermiittisiin ja definiittisiin matriiseihin ja muodostavat itsessäänkin mielenkiintoinen matriisien joukko.

3.1 Unitaariset matriisit

Tässä aliluvussa esitellään siis unitaariset matriisit, jotka ovat yleistetty muoto reaalisista ortogonaalisista matriiseista, joille $U^T U = I$. Kaikki aliluvussa käsiteltävät ominaisuudet siis toimivat myös ortogonaalisille matriiseille. Käsittelemme myös unitaarista similaarisuutta, joka on läheistä sukua similaarisuuden käsitteelle.

Määritelmä 3.1. Matriisia $U \in \mathbb{C}^{n\times n}$ sanotaan *unitaariseksi*, jos $U^*U = I$.

Jos matriisi $U \in \mathbb{R}^{n\times n}$, niin $U^* = U^T$, jolloin matriisia kutsutaan ortogonaaliseksi.

Esimerkki 3.1. Matriisi

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{3}} & \frac{i}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

on unitaarinen, koska

$$U^*U = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{-i}{\sqrt{2}} \\ \frac{-i}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{-i}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{3}} & \frac{i}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Seuraavassa lauseessa esittelemme joitakin unitaaristen matriisien tärkeimpiä ominaisuuksia, joita tarvitsemme myöhemmin tutkielmassa.

Lause 3.1. *Jos $U \in \mathbb{C}^{n\times n}$, niin seuraavat ominaisuudet ovat ekvivalentteja.*

(1) Matriisi U on unitaarinen,

(2) matriisi U on kääntyvä ja $U^{-1} = U^*$,

(3) $UU^* = I$,

(4) matriisi U^* on unitaarinen,

(5) matriisin U sarakkeet muodostavat ortonormaalin joukon,

(6) matriisin U rivit muodostavat ortonormaalin joukon,

(7) $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ kaikilla $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$.

Todistus. Oletetaan ensin, että U on unitaarinen matriisi, eli $U^*U = I$. Tällöin $U^* = U^{-1}$, koska U^{-1} on sellainen yksikäsitteinen matriisi, jolle $U^{-1}U = I$. Koska U^{-1} on olemassa, niin on matriisi U myös kääntyvä. Olemme siis todistaneet, että kohdasta (1) seuraa kohdan (2) väite.

Oletetaan nyt, että U on kääntyvä matriisi ja $U^{-1} = U^*$. Tällöin selvästi niin $UU^* = UU^{-1} = I$, joten kohdasta (2) seuraa kohdan (3) väite.

Oletetaan seuraavaksi, että $UU^* = I$. Koska $U = (U^*)^*$, niin $(U^*)^*U^* = I$, joten määritelmän 3.1 nojalla myös U^* on unitaarinen. Olemme siis todistaneet, että kohdasta (3) seuraa kohdan (4) väite.

Oletetaan U^* unitaariseksi ja todistetaan, että myös U on tällöin unitaarinen. Nyt määritelmän 3.1 nojalla $(U^*)^*U^* = I$. Koska neliömatriiseille $AB = I$, jos ja vain jos $BA = I$ ja $(U^*)^* = U$, niin $(U^*)^*U^* = UU^* = U^*U = I$, eli myös U on unitaarinen. Täten kohdasta (4) seuraa kohdan (1) väite.

Näin olemme siis todistaneet, että kohdat (1)-(4) ovat ekvivalentteja. Osoitetaan vielä, että kohdat (1) ja (5), (4) ja (6) sekä (1) ja (7) ovat ekvivalentteja keskenään, jonka jälkeen olemme todistaneet lauseen.

Oletetaan ensin, että vektorit $\boldsymbol{u_i} \in \mathbb{C}^n$, kun $i = 1, 2, \ldots, n$, muodostavat matriisin U sarakkeet, eli $U = \begin{bmatrix}\boldsymbol{u_1} & \cdots & \boldsymbol{u_n}\end{bmatrix}$, jolloin

$$U^*U = \begin{bmatrix}\boldsymbol{u_1^*} \\ \vdots \\ \boldsymbol{u_n^*}\end{bmatrix}\begin{bmatrix}\boldsymbol{u_1} & \cdots & \boldsymbol{u_n}\end{bmatrix} = \begin{bmatrix}\langle \boldsymbol{u_1}, \boldsymbol{u_1}\rangle & \cdots & \langle \boldsymbol{u_n}, \boldsymbol{u_1}\rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{u_1}, \boldsymbol{u_n}\rangle & \cdots & \langle \boldsymbol{u_n}, \boldsymbol{u_n}\rangle\end{bmatrix}.$$

Oletetaan, että U on unitaarinen, eli $U^*U = I$. Tällöin

$$\langle \boldsymbol{u_i}, \boldsymbol{u_j}\rangle = \begin{cases}1, & \text{kun } i = j \\ 0, & \text{kun } i \neq j.\end{cases}$$

Tämä tarkoittaa sitä, että vektorit $\boldsymbol{u_1}, \ldots, \boldsymbol{u_n}$ muodostavat ortonormaalin joukon. Kääntäen, jos oletetaan, että vektorit $\boldsymbol{u_1}, \ldots, \boldsymbol{u_n}$ muodostavat ortonormaalin joukon, niin saamme

$$U^*U = \begin{bmatrix}\langle \boldsymbol{u_1}, \boldsymbol{u_1}\rangle & \cdots & \langle \boldsymbol{u_n}, \boldsymbol{u_1}\rangle \\ \vdots & \ddots & \vdots \\ \langle \boldsymbol{u_1}, \boldsymbol{u_n}\rangle & \cdots & \langle \boldsymbol{u_n}, \boldsymbol{u_n}\rangle\end{bmatrix} = \begin{bmatrix}1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1\end{bmatrix} = I,$$

joten matriisi U on unitaarinen. Olemme siis todistaneet, että kohdat (1) ja (5) ovat ekvivalentteja. Kohdat (4) ja (6) todistettaisiin ekvivalenteiksi samalla tavalla kuin (1) ja (5), joten sivuutamme kyseisen todistuksen.

Osoitetaan vielä, että kohdat (1) ja (7) ovat ekvivalentteja. Oletetaan ensin, että $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ ja $U^*U = I$. Tällöin

$$\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle U^*U\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle.$$

Oletetaan sitten, että $\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ kaikilla $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, jolloin sisätulon laskusääntöjen nojalla saadaan $\langle (U^*U - I)\boldsymbol{x}, \boldsymbol{y}\rangle = 0$. Jos valitaan $\boldsymbol{y} = (U^*U - I)\boldsymbol{x}$, niin $\|(U^*U - I)\boldsymbol{x}\|^2 = 0$, joten $(U^*U - I)\boldsymbol{x} = \Theta$ kaikilla $\boldsymbol{x} \in \mathbb{C}^n$.Täten on oltava $(U^*U - I) = \mathbf{0}$, eli $U^*U = I$, joten matriisi U on unitaarinen. □

Seuraavassa lauseessa todistamme useamman unitaarisen matriisin tulon olevan myös unitaarinen. Lauseen muotoa on hieman muutettu kirjan [1] vastaavan lauseeseen verrattuna, sillä se todistaa vain kahden unitaarisen matriisin tulon unitaarisuuden.

Lause 3.2. *Olkoon matriisit $U_1, U_2, \ldots, U_m \in \mathbb{C}^{n\times n}$ unitaarisia. Tällöin myös matriisi $U_1U_2\cdots U_m \in \mathbb{C}^{n\times n}$ on unitaarinen.*

Todistus. Jos $U_1, U_2, \ldots, U_m \in \mathbb{C}^{n\times n}$ ovat unitaarisia, niin

$$\begin{aligned}(U_1U_2\cdots U_m)^*(U_1U_2\cdots, U_m) &= U_m^*\cdots U_2^*U_1^*U_1U_2\cdots U_m \\ &= U_m^*\cdots U_2^*IU_2\cdots U_m = U_m^*\cdots U_2^*U_2\cdots U_m = \cdots = I,\end{aligned}$$

joten myös $U_1U_2\cdots U_m$ on unitaarinen. □

Unitaarisen matriisien ominaisuuksia tarkastellessa huomasimme, että niille pätee $U^* = U^{-1}$. Voimmekin alkaa tarkastella sellaisia similaarisia matriiseja, jonka kääntyvä välittäjämatriisi on lisäksi myös unitaarinen.

Määritelmä 3.2. Matriiseja $A, B \in \mathbb{C}^{n\times n}$ kutsutaan keskenään *unitaarisesti similaarisiksi matriiseiksi*, jos on olemassa unitaarinen matriisi $U \in \mathbb{C}^{n\times n}$ siten, että

$$A = UBU^* = UBU^{-1}.$$

Jos lisäksi matriisi B on diagonaalimatriisi, niin matriisia A kutsutaan *unitaarisesti diagonalisoituvaksi.*

Selvästi unitaarisesti similaariset matriisit ovat myös similaarisia, joten kaikki similaarisuuteen liittyvät ominaisuudet pätevät myös unitaarisesti similaarisille matriiseille. Similaarisuuden tavoin myös unitaarinen similaarisuus on ekvivalenssirelaatio, kuten seuraava lause osoittaa.

Lause 3.3. *Unitaarinen similaarisuus on ekvivalenssirelaatio, eli*

(1) $A = UAU^*$,

(2) jos $A = UBU^*$, niin $B = VAV^*$,

(3) jos $A = UBU^*$ ja $B = VCV^*$, niin $A = WCW^*$, missä $W = UV$,

joillakin matriiseilla $A, B, C \in \mathbb{C}^{n\times n}$ ja unitaarisilla matriiseilla $U, V, W \in \mathbb{C}^{n\times n}$.

Todistus. Todistetaan ensin kohta (1). Valitaan $U = I$, joka on unitaarinen, jolloin selvästi $IAI^* = A$, joten unitaarinen similaarisuus on refleksiivinen relaatio.

Todistetaan seuraavaksi kohta (2). Olkoon $A = UBU^*$. Nyt jos kerromme kyseisen yhtälön vasemmalta matriisilla U^* ja oikealta matriisilla U, saamme yhtälön muotoon $U^*UBU^*U = B = U^*AU$. Koska lauseen 3.1 nojalla U^* on unitaarinen, jos U on unitaarinen, niin voimme merkitä $V = U^*$ ja saamme unitaarisesta similaarisuudesta myös symmetrisen relaation.

Todistetaan vielä kohta (3). Nyt jos $A = UBU^*$ ja $B = VCV^*$, saamme

$$A = UBU^* = U(VCV^*)U^* = (UV)C(UV)^*.$$

Koska lauseen 3.2 nojalla UV on unitaarinen, niin $W = UV$ ja unitaarinen similaarisuus on transitiivinen relaatio.

Koska unitaarinen similaarisuus on refleksiivinen, symmetrinen ja transitiivinen relaatio, niin se on ekvivalenssirelaatio. $\square$

Esimerkki 3.2. Vaikka kaikki unitaarisesti similaariset matriisit ovat myös similaarisia, niin toiseen suuntaan tämä ei kuitenkaa päde. Esimerkiksi

$$UBU^{-1} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ -2 & 0 \end{bmatrix} = A,$$

joten matriisit A ja B ovat similaarisia. Ne eivät ole kuitenkaan unitaarisesti similaarisia (todistus sivuutetaan).

Unitaarisen similaarisuuden kohdalla nousee esiin kysymys, kuinka yksinkertaisessa muodossa mielivaltainen matriisi voidaan esittää unitaarisen similaarisuuden avulla. Tähän kysymyksen vastauksen antaa Schurin hajotelmaksi kutsuttu lause, jonka mukaan jokainen neliömatriisi on unitaarisesti similaarinen jonkin yläkolmiomatriisin kanssa. Olemme käyttäneet tämän lauseen todistukseen lähteenä kirjan [1] algoritmityyppistä todistusta, mutta lause voitaisiin todistaa myös induktion avulla (katso esimerkiksi [2, s. 176-178]). Voisimme myös todistaa, että jokainen matriisi on unitaarisesti similaarinen jonkin alakolmiomatriisin kanssa, mutta sivuutamme kyseisen todistuksen. Yleisesti Schurin hajotelmasta käytetään seuraavaksi esiteltävää yläkolmiomuotoa.

Lause 3.4 (Schurin hajotelma)**.** *Olkoot $A \in \mathbb{C}^{n\times n}$ ja $\lambda_1, \dots, \lambda_n$ sen ominaisarvot. Tällöin on olemassa yläkolmiomatriisi $T \in \mathbb{C}^{n\times n}$, jonka diagonaalialkiot ovat matriisin A ominaisarvot, ja unitaarinen matriisi $U \in \mathbb{C}^{n\times n}$ siten, että*

$$A = UTU^*.$$

Toisin sanoen jokainen neliömatriisi A on unitaarisesti similaarinen jonkin yläkolmiomatriisin T kanssa, joka sisältää matriisin A ominaisarvot diagonaalialkioinaan.

Todistus. Merkitsemme diagonaalin yläpuolella olevia alkioita tai vektoreita merkillä $*$, jos niiden muodolla ei ole todistuksen kannalta olennaista merkitystä. Olkoot $\lambda_1, \dots, \lambda_n$ matriisin A ominaisarvot ja $\boldsymbol{x_1} \in \mathbb{C}^n$ ominaisarvoon λ_1 liittyvä ominaisvektori, jolle $\|\boldsymbol{x_1}\| = 1$. Täydennetään joukko $\{\boldsymbol{x_1}\}$ vektoriavaruuden $\mathbb{C}^n$ kannaksi ja muodostetaan Gram-Schmidtin ortogonalisointialgoritmin avulla tästä kannasta vektoriavaruuden $\mathbb{C}^n$ ortonormaali kanta $\beta_1 = \{\boldsymbol{x_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n}\}$. Järjestetään kannan β_1 vektorit matriisiksi

$$U_1 = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{y_2} & \dots & \boldsymbol{y_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} & Y \end{bmatrix}.$$

Koska $A\boldsymbol{x_1} = \lambda_1 \boldsymbol{x_1}$ ja $\langle \boldsymbol{x_1}, \boldsymbol{y_j} \rangle = \boldsymbol{y_j}^* \boldsymbol{x_1} = 0$, kun $j = 2, \dots, n$, saamme

$$\begin{aligned} U_1^* A U_1 &= \begin{bmatrix} \boldsymbol{x_1}^* \\ Y^* \end{bmatrix} A \begin{bmatrix} \boldsymbol{x_1} & Y \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1}^* A \boldsymbol{x_1} & \boldsymbol{x_1}^* A Y \\ Y^* A \boldsymbol{x_1} & Y^* A Y \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{x_1}^* \lambda_1 \boldsymbol{x_1} & * \\ Y^* \lambda_1 \boldsymbol{x_1} & A_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 \boldsymbol{x_1}^* \boldsymbol{x_1} & * \\ \lambda_1 Y^* \boldsymbol{x_1} & A_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & * \\ \Theta & A_1 \end{bmatrix} = C, \end{aligned}$$

jossa $A_1 = Y^* A Y$. Nyt matriisit A ja C ovat unitaarisesti similaarisina myös similaariset, joten niillä on samat ominaisarvot lauseen 2.5 nojalla. Tällöin siis matriisin $A_1 \in \mathbb{C}^{(n-1)\times(n-1)}$ ominaisarvot ovat $\lambda_2, \dots, \lambda_n$. Olkoon nyt $\boldsymbol{x_2} \in \mathbb{C}^{(n-1)}$ matriisin A_1 ominaisarvoa λ_2 vastaava ominaisvektori. Tehdään tälle ominaisvektorille samat operaatiot kuin aikaisemmin vektorille $\boldsymbol{x_1}$, jolloin saadaan vektoriavaruuden $\mathbb{C}^{(n-1)}$ kanta $\beta = \{\boldsymbol{x_2}, \boldsymbol{z_3}, \dots, \boldsymbol{z_n}\}$. Määritellään unitaarinen matriisi $U_2 \in \mathbb{C}^{(n-1)\times(n-1)}$ siten, että

$$U_2 = \begin{bmatrix} \boldsymbol{x_2} & \boldsymbol{z_3} & \cdots & \boldsymbol{z_n} \end{bmatrix},$$

jolloin

$$U_2^* A_1 U_2 = \begin{bmatrix} \lambda_2 & * \\ \Theta & A_2 \end{bmatrix}.$$

Muodostetaan sitten matriisi

$$V_2 = \begin{bmatrix} 1 & \Theta \\ \Theta & U_2 \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

Huomataan, että matriisit V_2 ja $U_1 V_2$ ovat unitaarisia. Nyt saadaan matriisi $V_2^* U_1^* A U_1 V_2$ muotoon

$$V_2^* U_1^* A U_1 V_2 = \begin{bmatrix} \lambda_1 & * \\ \Theta & U_2^* A_1 U_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & * & * \\ 0 & \lambda_2 & * \\ \Theta & \Theta & A_2 \end{bmatrix}.$$

Jatkamalla algoritmia unitaarisille matriiseille U_i, jossa $i = 1, 2, \dots, n-1$, ja V_i, jossa $i = 2, \dots, n-1$, ja merkitsemällä $U = U_1 V_2 \cdots V_{n-1}$ saamme matriisin

$$V_{n-1}^* \cdots V_2^* U_1^* A U_1 V_2 \cdots V_{n-1} = U^* A U = T,$$

jossa $T \in \mathbb{C}^{n\times n}$ on haluttua muotoa oleva yläkolmiomatriisi ja matriisi U on unitaarinen. Nyt siis $U^*AU = T$, josta saamme $A = UTU^*$, eli matriisi A ja yläkolmiomatriisi T ovat unitaarisesti similaarisia. $\square$

Edeltäneessä Schurin hajotelman todistuksessa osoitimme, että jokainen neliömatriisi on unitaarisesti similaarinen jonkun yläkolmiomatriisin kanssa. Emme kuitenkaan osoittaneet, että kyseinen yläkolmiomatriisi olisi yksikäsitteinen. Näin ei olekaan, sillä todistusta tarkastellessa voimme huomata, että voimme valita ominaisarvot diagonaalille missä järjestyksessä tahansa.

Esimerkki 3.3. Tarkastellaan matriisia

$$A = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix},$$

joka on selvästi unitaarisesti similaarinen itsensä ja täten jonkin yläkolmiomatriisin kanssa. Toisaalta se on unitaarisesti similaarinen myös yläkolmiomatriisin

$$T = \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}$$

kanssa, sillä

$$UTU^* = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} = A.$$

3.2 Normaalit matriisit

Tässä aliluvussa esittelemme normaaleja matriiseja ja niiden ominaisuuksia. Olemme määritelleet normaalit matriisit erilailla kuin käyttämämme lähdekirja [3]. Tämä johtuu lähinnä kahdesta syystä. Ensinnäkin normaalit matriisit määritellään yleensä kirjallisuudessa esittämällämme tavalla ja toisekseen määritelmämme on tyyliltään sekä unitaaristen että hermiittisten matriisien määritelmien kanssa yhtäläinen.

Määritelmä 3.3. Matriisia $A \in \mathbb{C}^{n\times n}$ kutsutaan *normaaliksi matriisiksi*, jos $A^*A = AA^*$, eli jos A kommutoi sen konjugaattitranspoosin kanssa.

Tarkastelemalla normaalien matriisien määritelmää voimme huomata helposti, että kaikki unitaariset matriisit ovat myös normaaleja matriiseja, sillä $U^*U = I = UU^*$. Kaikki normaalit matriisit eivät kuitenkaan ole unitaarisia, kuten voimme seuraavasta esimerkistä huomata.

Esimerkki 3.4. Tarkastellaan matriisia

$$A = \begin{bmatrix} i & 0 & 1 \\ 1 & i & 0 \\ 0 & 1 & i \end{bmatrix}.$$

Koska

$$A^*A = \begin{bmatrix} -i & 1 & 0 \\ 0 & -i & 1 \\ 1 & 0 & -i \end{bmatrix} \begin{bmatrix} i & 0 & 1 \\ 1 & i & 0 \\ 0 & 1 & i \end{bmatrix} = \begin{bmatrix} 2 & i & -i \\ -i & 2 & i \\ i & -i & 2 \end{bmatrix}$$

ja

$$AA^* = \begin{bmatrix} i & 0 & 1 \\ 1 & i & 0 \\ 0 & 1 & i \end{bmatrix} \begin{bmatrix} -i & 1 & 0 \\ 0 & -i & 1 \\ 1 & 0 & -i \end{bmatrix} = \begin{bmatrix} 2 & i & -i \\ -i & 2 & i \\ i & -i & 2 \end{bmatrix},$$

niin matriisi A on normaali. Matriisi A ei kuitenkaan ole unitaarinen (eikä hermiittinen).

Schurin hajotelma aikaisemmin osoitti, että mikä tahansa neliömatriisi on unitaarisesti similaarinen jonkun yläkolmiomatriisin kanssa. Normaaleista matriiseista voimme unitaarisen similaarisuuden avulla sanoa vielä enemmän, kuten seuraava lause osoittaa.

Lause 3.5. *Matriisi $A \in \mathbb{C}^{n\times n}$ on normaali, jos ja vain jos se on unitaarisesti similaarinen sen ominaisarvoista $\lambda_1, \lambda_2, \ldots, \lambda_n$ muodostetun diagonaalimatriisin $D \in \mathbb{C}^{n\times n}$ kanssa, eli*

$$A = UDU^*,$$

jossa $D = diag\ (\lambda_i)$, kun $i = 1, 2, \ldots, n$, ja $U \in \mathbb{C}^{n\times n}$ on unitaarinen.

Todistus. Oletetaan ensin, että $A = UDU^*$, jossa D on matriisin A ominaisarvoista muodostettu diagonaalimatriisi ja U unitaarinen. Tällöin

$$\begin{aligned} AA^* &= (UDU^*)(UDU^*)^* = UDU^*UD^*U^* = UDD^*U^* \\ &= UD^*DU^* = UD^*U^*UDU^* = (UDU^*)^*(UDU^*) = A^*A, \end{aligned}$$

joten matriisi A on normaali.

Oletetaan nyt, että A on normaali, eli $AA^* = A^*A$. Nyt lauseen 3.4 nojalla on olemassa unitaarinen matriisi $U \in \mathbb{C}^{n\times n}$ ja yläkolmiomatriisi $T \in \mathbb{C}^{n\times n}$ siten, että $A = UTU^*$. Koska $AA^* = UTU^*UT^*U^* = UTT^*U^*$ ja $A^*A = UT^*U^*UTU^* = UT^*TU^*$, niin selvästi tulee olla myös

$$TT^* = T^*T. \tag{3.1}$$

Nyt yhtälön (3.1) ensimmäisen rivin ensimmäiselle alkiolle saadaan

$$|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 = |t_{11}|^2,$$

josta saamme $|t_{12}|^2 + \cdots + |t_{1n}|^2 = 0$, joten $t_{1i} = 0$, kun $i = 2, 3, \ldots, n$. Yhtälön (3.1) toisen rivin toiselle alkiolle taasen saadaan

$$|t_{22}|^2 + |t_{23}|^2 + \cdots + |t_{2n}|^2 = |t_{12}|^2 + |t_{22}|^2,$$

josta saamme $|t_{23}|^2 + \cdots + |t_{2n}|^2 = 0$, joten $t_{2i} = 0$, kun $i = 3, \ldots, n$. Näin jatkamalla saamme $t_{jk} = 0$, kun $j \neq k$, $j = 1, 2, \ldots, n$ ja $k = 1, 2, \ldots, n$, joten matriisin T on diagonaalimatriisi. Lisäksi lauseen 3.4 nojalla matriisin T diagonaalialkiot ovat matriisin A ominaisarvot, joten olemme todistaneet lauseen. □

Seuraavaksi esitämme tärkeän lauseen normaalien matriisien ominaisvektoreista. Kirjassa [3] kyseistä ominaisuutta on käytetty normaalien matriisien määrittelyyn.

Lause 3.6. *Olkoon matriisi A normaali. Tällöin sillä on n kappaletta ortonormaaleja ominaisvektoreita.*

Todistus. Koska matriisi A on normaali, niin se voidaan kirjoittaa edellisen lauseen avulla muodossa $A = UDU^*$, joten se on diagonalisoituva. Tällöin lauseen 2.6 nojalla huomaamme, että matriisi U koostuu matriisin A ominaisvektoreista. Koska matriisin U sarakkeet ovat ortonormaaleja, niin matriisilla A on täten aina n kappaletta ortonormaaleja ominaisvektoreita. □

Täten siis huomaamme, että normaaleilla matriiseilla on aina olemassa ortonormaali ominaisvektorikanta. Käytämme tätä huomiota usein tutkielman myöhemmissä osissa, kuten myös normaalien matriisien hajotelmaa $A = UDU^*$, joka on erittäin kätevä muoto useissa todistuksissa.

Lause 3.7. *Olkoon matriisi $A \in \mathbb{C}^{n \times n}$ normaali, olkoot $\lambda_1, \lambda_2, \ldots, \lambda_n$ sen ominaisarvot ja olkoot $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$ niihin liittyvät ortonormaalit ominaisvektorit. Tällöin matriisin A^* ominaisarvot ovat $\overline{\lambda_1}, \overline{\lambda_2}, \ldots, \overline{\lambda_n}$ ja ominaisarvoihin liittyvät ortonormaalit ominaisvektorit ovat $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$.*

Todistus. Koska matriisi A on normaali, niin se voidaan kirjottaa muotoon $A = UDU^*$, jossa $D = \operatorname{diag}(\lambda_i) \in \mathbb{C}^{n \times n}$, kun $i = 1, 2, \ldots, n$ ja $U \in \mathbb{C}^{n \times n}$ on muodostettu ortonormaaleista ominaisvektoreista $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$. Nyt saamme

$$A^* = (UDU^*)^* = U\overline{D}U^*.$$

Koska $\overline{D}$ sisältää nyt diagonaalillaan matriisi A^* ominaisarvot, niin ne ovat muotoa $\overline{\lambda_i}$, kun $i = 1, 2, \ldots, n$. Nyt myös U sisältää matriisin A^* ominaisvektorit, jotka ovat siis samat kuin matriisilla A. □

Luvun alussa huomasimme, että kaikki unitaariset matriisit ovat aina myös normaaleja. Tästä nousee esille kysymys, millä lisäehdoilla saamme normaalista matriisista unitaarisen. Seuraava lause antaa tähän vastauksen.

Lause 3.8. *Olkoon matriisi $A \in \mathbb{C}^{n \times n}$ normaali. Tällöin A on unitaarinen, jos ja vain jos sen kaikki ominaisarvot sijaitsevat yksikköympyrällä.*

Todistus. Olkoot matriisin A ominaisarvot $\lambda_1, \lambda_2, \ldots, \lambda_n$ ja olkoot niihin liittyvät ortonormaalit ominaisvektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$. Tällöin $A\boldsymbol{x_i} = \lambda_i \boldsymbol{x_i}$ ja $\langle \boldsymbol{x_i}, \boldsymbol{x_i} \rangle = 1$, kun $i = 1, 2, \ldots, n$.

Oletetaan ensin, että matriisi A on unitaarinen. Tällöin

$$\langle A\boldsymbol{x_i}, A\boldsymbol{x_i} \rangle = \langle A^* A\boldsymbol{x_i}, \boldsymbol{x_i} \rangle = \langle \boldsymbol{x_i}, \boldsymbol{x_i} \rangle = 1.$$

Toisaalta

$$\langle A\boldsymbol{x_i}, A\boldsymbol{x_i} \rangle = \langle \lambda_i \boldsymbol{x_i}, \lambda_i \boldsymbol{x_i} \rangle = |\lambda_i|^2 \langle \boldsymbol{x_i}, \boldsymbol{x_i} \rangle = |\lambda_i|^2 ,$$

josta saamme $|\lambda_i|^2 = 1$, joten matriisin A kaikki ominaisarvot sijaitsevat yksikköympyrällä.

Oletetaan sitten, että normaalin matriisin A kaikki ominaisarvot sijaitsevat yksikköympyrällä. Koska vektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$ muodostavat vektoriavaruuden $\mathbb{C}^n$ kannan, niin voimme mille tahansa vektorille $\boldsymbol{x} \in \mathbb{C}^n$ kirjoittaa

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x_i},$$

jonka avulla saamme

$$A^* A\boldsymbol{x} = A^* \sum_{i=1}^{n} \alpha_i A\boldsymbol{x_i} = A^* \sum_{i=1}^{n} \alpha_i \lambda_i \boldsymbol{x_i}.$$

Lauseen 3.7 nojalla $A^* \boldsymbol{x_i} = \bar{\lambda}_i \boldsymbol{x_i}$, joten saamme

$$A^* A\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \lambda_i A^* \boldsymbol{x_i} = \sum_{i=1}^{n} \alpha_i |\lambda_i|^2 \boldsymbol{x_i} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x_i} = \boldsymbol{x}.$$

Nyt $A^* A\boldsymbol{x} = \boldsymbol{x}$ kaikilla $\boldsymbol{x} \in \mathbb{C}^n$, joten $A^* A = I = AA^*$, eli matriisi A on unitaarinen. $\square$

4 Hermiittiset ja positiivisesti (semi)definiitit matriisit

Tässä luvussa käsittelemme ensin hermiittisiä matriiseja, jonka jälkeen käsittelemme niihin liittyviä definiittejä matriiseja. Hermiittisten matriisien käsittely rajoitetaan lähinnä johdattelemaan lukija definiitteihin matriiseihin. Lisäksi hermiittisistä matriiseista käydään läpi joitakin myöhemmin tarvittavia ominaisuuksia. Definiittejä matriiseja käsitellään hieman syvällisemmin kuin hermiittisiä matriiseja, koska ne ovat tärkeässä osassa luvussa 5. Johdattalemmekin definiittien matriisien avulla lukijan singulaariarvojen käsitteeseen, jota käsitellään luvussa 5.1.

4.1 Hermiittiset matriisit

Reaaliarvoista matriisia kutsutaan symmetriseksi, jos $A = A^T$. Tässä aliluvussa tutkitaan kompleksiarvoisten matriisen vastaavaa ominaisuutta.

Määritelmä 4.1. Matriisia $A \in \mathbb{C}^{n\times n}$ kutsutaan *hermiittiseksi matriisiksi*, jos pätee $A = A^*$. Jos $A = -A^*$, niin matriisia A kutsutaan *vinohermiittiseksi matriisiksi*.

Määritelmän perusteella voimme helposti huomata, että hermiittiset ja vinohermiittiset matriisit ovat aina myös normaaleja, sillä hermiittisille matriiseille pätee $A^*A = A^2 = AA^*$ ja vinohermiittisille $A^*A = -A^2 = AA^*$. Normaalit matriisit eivät kuitenkaan aina ole hermiittisiä tai vinohermiittisiä, kuten voimme huomata edellisen luvun esimerkin 3.4 tapauksessa. Joillakin lisäehdoilla normaalista matriisista tosin saadaan hermiittinen matriisi. Tätä asiaa käsitellään tarkemmin lauseessa 4.6. Seuraavissa kahdessa lauseessa käsittelemme joitain hermiittisyyteen liittyviä ominaisuuksia.

Lause 4.1. *Jos $A \in \mathbb{C}^{n\times n}$, niin matriisit $A + A^*, AA^*, A^*A$ ovat hermiittisiä ja matriisi $A - A^*$ on vinohermiittinen.*

Todistus. Koska

$$(A + A^*)^* = A^* + (A^*)^* = A + A^*,$$

niin $A + A^*$ on hermiittinen. Nyt $(AA^*)^* = (A^*)^*A^* = AA^*$ ja vastaavasti $(A^*A)^* = A^*(A^*)^* = A^*A$, joten myös AA^* ja A^*A ovat hermiittisiä. Koska

$$(A - A^*)^* = A^* - A = -(A - A^*),$$

niin matriisi $A - A^*$ on vinohermiittinen. □

Lause 4.2. *Seuraavassa luetellaan joitakin hermiittisiin ja vinohermiittisiin matriiseihin liittyviä ominaisuuksia.*

(1) Jos $A \in \mathbb{C}^{n\times n}$ on hermiittinen, niin myös A^k on hermiittinen kaikilla $k \in \mathbb{Z}_+$.

(2) Jos $A \in \mathbb{C}^{n\times n}$ on hermiittinen ja kääntyvä, niin A^{-1} on hermiittinen.

(3) Jos matriisit $A, B \in \mathbb{C}^{n\times n}$ ovat (vino)hermiittisiä ja skalaarit $a, b \in \mathbb{R}$, niin myös $aA + bB$ on (vino)hermiittinen.

(4) Jos $A \in \mathbb{C}^{n\times n}$ on (vino)hermiittinen, niin iA on vinohermiittinen (hermiittinen).

Todistus. Koska

$$(A^k)^* = (\underbrace{A\cdots A}_{\text{k kpl}})^* = \underbrace{A^*\cdots A^*}_{\text{k kpl}} = (A^*)^k = A^k,$$

niin A^k on hermiittinen ja kohta (1) on voimassa.

Kohta (2) pätee triviaalisti lauseen 2.1, sillä $(A^{-1})^* = (A^*)^{-1} = A^{-1}$.

Todistetaan kohta (3) kahdessa osassa, ensin hermiittisyyden osalta ja sitten vinohermiittisyyden osalta. Nyt jos A ja B ovat hermiittisiä, niin

$$(aA + bB)^* = (aA)^* + (bB)^* = A^*a^* + B^*b^* = aA + bB,$$

joten $aA + bB$ on hermiittinen. Jos taas A ja B ovat vinohermiittisiä, niin

$$(aA + bB)^* = A^*a^* + B^*b^* = -aA - bB = -(aA + bB),$$

joten tällöin $aA + bB$ on vinohermiittinen.

Todistetaan kohta (4) taas kahdessa osassa. Jos A on hermiittinen, niin $(iA)^* = A^*i^* = -iA$, joten iA on vinohermiittinen. Jos A on vinohermiittinen, niin $(iA)^* = A^*i^* = iA$, joten tällöin iA on hermiittinen. □

Tarkastellaan mielivaltaista matriisia $A \in \mathbb{C}^{n\times n}$. On helppo huomata, että A voidaan esittää aina muodossa

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*) = H(A) + S(A).$$

Tällöin lauseketta $H(A) = \frac{1}{2}(A + A^*)$ kutsutaan *matriisin A hermiittiseksi osaksi* ja lauseketta $S(A) = \frac{1}{2}(A - A^*)$ kutsutaan *matriisin A vinohermiittiseksi osaksi.* Tässä saamme selvän analogian imaginäärilukujen ja neliömatriisien välille, jota emme tosin käsittele tässä tutkielmassa seuraavaa lausetta enempää. Jos samaistamme joukon $\mathbb{C}^{n\times n}$ joukon $\mathbb{C}$ kanssa, niin hermiittiset matriisit ovat analogisia reaalilukujen joukon $\mathbb{R}$ kanssa. Seuraava lause vahvistaa tätä analogiaa.

Lause 4.3. *Jokainen matriisi $A \in \mathbb{C}^{n\times n}$ voidaan esittää yksikäsitteisesti muodossa $A = S + iT$, jossa S ja T ovat molemmat hermiittisiä. Matriisi A voidaan myös esittää yksikäsitteisesti muodossa $A = B + C$, jossa B on hermiittinen ja C vinohermiittinen matriisi.*

Todistus. Osoitetaan ensin, että matriisi $A \in \mathbb{C}^{n\times n}$ voidaan esittää muodossa $A = \frac{1}{2}(A + A^*) + i(\frac{-i}{2}(A - A^*))$. Nyt nimittäin

$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*) = \frac{1}{2}(A + A^*) + i(\frac{-i}{2}(A - A^*)),$$

joten voimme esittää matriisin A kyseisessä muodossa.

Merkitään $S = \frac{1}{2}(A + A^*)$ ja $T = \frac{-i}{2}(A - A^*)$ ja osoitetaan seuraavaksi, että S ja T ovat hermiittisiä. Lauseen 4.1 nojalla $A + A^*$ on hermiittinen ja lauseen 4.2 kohdan (3) nojalla $\frac{1}{2}(A+A^*)$ on hermiittinen, joten matriisi S on hermiittinen. Lauseen 4.1 nojalla $A - A^*$ on vinohermiittinen, joten lauseen 4.2 kohdan (4) nojalla $i(A - A^*)$ on hermiittinen. Tällöin saman lauseen kohdan (3) nojalla $-\frac{i}{2}(A - A^*)$ on hermiittinen, joten näin on todistettu, että myös T on hermiittinen. Osoittaaksemme esityksen yksikäsitteisyyden, teemme vastaoletuksen, että on olemassa toiset hermiittiset matriisit E ja F, joilla $A = E + iF$. Tällöin

$$2S = A + A^* = (E + iF) + (E + iF)^* = E + iF + E - iF = 2E,$$

joten $S = E$. Vastaavasti

$$2iT = A - A^* = (E + iF) - (E + iF)^* = E + iF - E + iF = 2iF,$$

joten $T = F$. Esitysmuoto $A = S + iT$ on siis yksikäsitteinen

Esitysmuotoon $A = B+C$ liittyvät tarkastelut samanlaiset kuin muotoon $A = S+iT$ liittyvät tarkastelut, koska voimme helposti huomata, että $B = S$ ja $C = iT$. Tällöin tarvitsee vain todistaa matriisin C vinohermiittisyys, joka seuraa suoraan lauseen 4.2 kohdasta (4). □

Seuraavan kolmen lauseen avulla todistamme, että normaali matriisi on hermiittinen, jos ja vain jos sen ominaisarvot ovat positiivisia reaalilukuja. Toinen mielenkiintoinen tulos näistä lauseista on se, että hermiittisten matriisien ominaisarvot ovat aina reaalilukuja. Tätä ominaisuutta käsittelemme lisää aliluvussa 4.2.

Lause 4.4. *Olkoon matriisi $A \in \mathbb{C}^{n\times n}$ hermiittinen. Tällöin $\langle A\boldsymbol{x}, \boldsymbol{x}\rangle \in \mathbb{R}$ jokaisella $\boldsymbol{x} \in \mathbb{C}^n$.*

Todistus. Merkitään $\langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \boldsymbol{x}^*A\boldsymbol{x} = z \in \mathbb{C}$. Koska

$$z = \boldsymbol{x}^*A\boldsymbol{x} = \boldsymbol{x}^*A^*\boldsymbol{x} = (\boldsymbol{x}^*A\boldsymbol{x})^* = z^* = \overline{z},$$

niin $z = \langle A\boldsymbol{x}, \boldsymbol{x}\rangle \in \mathbb{R}$. □

Lause 4.5. *Olkoon matriisi $A \in \mathbb{C}^{n\times n}$ hermiittinen. Tällöin matriisin A kaikki ominaisarvot ovat reaalilukuja.*

Todistus. Olkoot $\lambda_1, \lambda_2, \ldots, \lambda_n$ matriisin A ominaisarvot ja olkoot vektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$ niihin liittyvät ortonormaalit ominaisvektorit. Koska $A\boldsymbol{x_i} = \lambda_i \boldsymbol{x_i}$ ja $\langle \boldsymbol{x_i}, \boldsymbol{x_i} \rangle = \boldsymbol{x_i}^* \boldsymbol{x_i} = 1$, kun $i = 1, 2, \ldots, n$, saamme lauseen 4.4 avulla

$$\lambda_i = \lambda_i \boldsymbol{x_i}^* \boldsymbol{x_i} = \boldsymbol{x_i}^* \lambda_i \boldsymbol{x_i} = \boldsymbol{x_i}^* A \boldsymbol{x_i} = \langle A\boldsymbol{x}, \boldsymbol{x} \rangle \in \mathbb{R},$$

joten kaikki hermiittisen matriisin ominaisarvot ovat reaalilukuja. □

Lause 4.6. *Olkoon matriisi $A \in \mathbb{C}^{n \times n}$ normaali. Tällöin matriisi A on hermiittinen, jos ja vain jos sen ominaisarvot ovat reaalilukuja.*

Todistus. Lauseessa 4.5 todistimme jo, että jos A on hermiittinen, niin sen ominaisarvot ovat reaalilukuja.

Oletetaan, että matriisi A on normaali ja että sen kaikki ominaisarvot ovat reaalilukuja. Lauseen 3.5 nojalla voimme kirjoittaa $A = UDU^*$, jossa matriisi $U \in \mathbb{C}^{n \times n}$ on unitaarinen ja $D \in \mathbb{C}^{n \times n}$ on diagonaalimatriisi, jonka diagonaalialkiot ovat matriisin A ominaisarvot. Tällöin saamme

$$A^* = (UDU^*)^* = U\overline{D}U^* = UDU^* = A,$$

joten matriisi A on hermiittinen. □

4.2 Positiivisesti (semi)definiitit matriisit

Huomasimme edellisessä kappaleessa, että hermiittisten matriisien ominaisarvot ovat aina reaalilukuja. On luonnollista alkaa tutkia minkälaisia ominaisuuksia hermiittinen matriisi saa, jos sen ominaisarvot ovat pelkästään positiivisia tai ei-negatiivisia.

Määritelmä 4.2. Olkoon matriisi $A \in \mathbb{C}^{n \times n}$ hermiittinen ja olkoot sen ominaisarvot $\lambda_1, \ldots, \lambda_n$. Tällöin matriisia A kutsutaan *positiivisesti definiitiksi*, jos sen kaikki ominaisarvot ovat positiivisia reaalilukuja, eli $\lambda_i > 0$, kun $i = 1, 2, \ldots, n$. Jos taas matriisin A ominaisarvot ovat kaikki ei-negatiivisia reaalilukuja, eli $\lambda_i \geq 0$, kun $i = 1, 2, \ldots, n$, kutsutaan matriisia *positiivisesti semidefiniitiksi*. Vastaavasti voidaan määritellä *negatiivisesti (semi)definiitit matriisit.*

On luonnollisempaa tutkia vain positiivisesti (semi)definiittejä matriiseja, sillä niillä on enemmän käyttöä myöhemmin tutkielmassa. Kuitenkin kaikki ominaisuudet, joita käsittelemme tässä luvussa, pätevät tietyillä muutoksilla myös negatiivisesti (semi)definiiteille matriiseille. Tämä johtuu siitä, että tutkiessamme negatiivisesti (semi)definiittiä matriisia A voimme tutkia matriisia $-A$, joka on positiivisesti (semi)definiitti ja johon siten pätee kaikki niiden ominaisuudet.

Lause 4.7. *Olkoon $A \in \mathbb{C}^{n \times n}$ positiivisesti semidefiniitti matriisi ja olkoon rank $A = r$. Tällöin matriisilla A on r kappaletta positiivista ominaisarvoa.*

Todistus. Koska matriisi A on hermiittinen ja täten myös normaali, voidaan se lauseen 3.5 nojalla kirjottaa muotoon $A = UDU^*$, jossa matriisi $U \in \mathbb{C}^{n\times n}$ on unitaarinen ja matriisi $D \in \mathbb{C}^{n\times n}$ on diagonaalimatriisi, jonka diagonaalialkiot ovat matriisin A ominaisarvot. Nyt matriisit A ja D ovat unitaarisesti similaarisia, eli myös similaarisia, joten $\operatorname{rank} A = \operatorname{rank} D = r$. Tällöin siis matriisilla D on r kappaletta sellaisia rivejä, jotka eivät ole nollarivejä, joten nollasta eroavia diagonaalialkioita myös on r kappaletta. Koska diagonaalialkiot ovat matriisin A ominaisarvot ja matriisi A on oletuksen nojalla positiivisesti semidefiniitti, niin nämä r kappaletta nollasta eroavia ominaisarvoja ovat positiivisia. □

Seuraavassa lauseessa osoitamme, että positiivisesti (semi)definiitille matriisille löytyy aina sitä vastaava positiivisesti semidefiniitti neliöjuurimatriisi. Tämä tieto osoittautuu myöhemmin tärkeäksi käsitellessämme singulaariarvoja.

Lause 4.8. *Matriisi $A \in \mathbb{C}^{n\times n}$ on positiivisesti (semi)definiitti, jos ja vain jos sillä on olemassa yksikäsitteinen positiivisesti (semi)definiitti neliöjuurimatriisi $A_0 \in \mathbb{C}^{n\times n}$, jolle $A_0^2 = A$. Lisäksi* $\operatorname{rank} A = \operatorname{rank} A_0$.

Todistus. Oletetaan, että matriisi A on positiivisesti definiitti ja $\lambda_1, \dots, \lambda_n$ ovat sen ominaisarvot, jolloin määritelmän 4.2 nojalla ne ovat positiivisia reaalilukuja. Täten voimme määritellä matriisin $D_0 = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$. Koska matriisi A on normaali, on olemassa unitaarinen matriisi $U \in \mathbb{C}^{n\times n}$ ja diagonaalimatriisi $D \in \mathbb{C}^{n\times n}$ siten, että $A = UDU^*$ ja $D_0^2 = D$. Nyt haluttu neliöjuurimatriisi on $A_0 = UD_0U^*$, sillä

$$A_0^2 = (UD_0U^*)^2 = UD_0^2U^* = UDU^* = A.$$

Matriisi A_0 on selvästi positiivisesti definiitti, johtuen matriisin D_0 määrittelystä. Osoitetaan seuraavaksi, että positiivisesti definiitti neliöjuurimatriisi A_0 on yksikäsitteinen. Oletetaan, että on olemassa positiivisesti definiitti matriisi $A_1 \in \mathbb{C}^{n\times n}$, jolle pätee $A_1^2 = A$. Nyt

$$A_1^2 = (UD_1U^*)^2 = UD_1^2U^* = UDU^* = A,$$

jossa $D_1 = \operatorname{diag}(\pm\sqrt{\lambda_1}, \dots, \pm\sqrt{\lambda_n})$. Koska A_1 on positiivisesti definiitti, niin sen ominaisarvot ovat positiivisia, joten tulee olla $D_1 = \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$. Nyt $D_1 = D_0$, joten $UD_1U^* = UD_0U^*$, eli $A_1 = A_0$. Täten siis neliöjuurimatriisi A_0 on yksikäsitteinen.

Oletetaan seuraavaksi, että $A_0^2 = A$ ja A_0 on positiivisesti definiitti. Tällöin matriisin A ominaisarvot ovat matriisin A_0 ominaisarvojen neliöitä, joten ne kaikki ovat ei-negatiivisia, eli matriisi A on myös positiivisesti definiitti.

Osoitetaan vielä, että $\operatorname{rank} A_0 = \operatorname{rank} A$. Koska matriisin D_0 diagonaalialkiot (eli matriisin A_0 ominaisarvot) ovat matriisin D neliöjuuria, niin matriiseilla D_0 ja D on yhtä monta positiivista diagonaalialkiota. Nyt selvästi $\operatorname{rank} A_0 = \operatorname{rank} D_0 = \operatorname{rank} D = \operatorname{rank} A$. □

Todistuksesta huomaamme, että λ on matriisin A_0 ominaisarvo, jos ja vain jos λ^2 on matriisin A ominaisarvo, ja että matriisien A ja A_0 ominaisavaruudet vastaavat toisiaan. Jatkossa käytämme neliöjuurimatriisista A_0 tutumpaa merkintää $\sqrt{A}$.

Seuraavaksi esitämme lauseen, jota joissakin lineaarialgebraa käsittelevissä kirjoissa käytetään definiittien matriisien määrittelemisessä. Vaikka tätä lausetta ei käytetä myöhemmin tutkielmassa, on se kuitenkin joissakin tapauksissa hyödyllinen työväline definiittien matriisien käsittelyssä.

Lause 4.9. *Olkoon $A \in \mathbb{C}^{n\times n}$ hermiittinen matriisi. Tällöin A on positiivisesti (semi)definiitti, jos ja vain jos $\langle A\boldsymbol{x}, \boldsymbol{x}\rangle$ on positiivinen (ei negatiivinen) kaikilla $\boldsymbol{x} \in \mathbb{C}^n$.*

Todistus. Olkoot vektorit $\boldsymbol{u_1}, \dots, \boldsymbol{u_n}$ matriisin A ominaisarvoja $\lambda_1, \dots, \lambda_n$ vastaavat ominaisvektorit, jotka muodostavat vektoriavaruuden $\mathbb{C}^n$ ortonormaalin kannan. Oletetaan, että $\langle A\boldsymbol{x}, \boldsymbol{x}\rangle$ on positiiviinen kaikilla $\boldsymbol{x} \in \mathbb{C}^n$. Tällöin

$$0 < \langle A\boldsymbol{u_i}, \boldsymbol{u_i}\rangle = \langle \lambda_i\boldsymbol{u_i}, \boldsymbol{u_i}\rangle = \lambda_i \langle \boldsymbol{u_i}, \boldsymbol{u_i}\rangle = \lambda_i,$$

kun $i = 1, 2, \dots, n$, joten matriisi A on positiivisesti definiitti.

Oletetaan seuraavaksi, että A on positiivisesti definiitti matriisi ja osoitetaan, että tällöin $\langle A\boldsymbol{x}, \boldsymbol{x}\rangle$ on positiivinen kaikilla $\boldsymbol{x} \in \mathbb{C}^n$. Lauseen 4.8 nojalla $A = (\sqrt{A})^2$, joten

$$\langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \left\langle (\sqrt{A})^2\boldsymbol{x}, \boldsymbol{x}\right\rangle = \left\langle \sqrt{A}\boldsymbol{x}, (\sqrt{A})^*\boldsymbol{x}\right\rangle = \left\langle \sqrt{A}\boldsymbol{x}, \sqrt{A}\boldsymbol{x}\right\rangle > 0,$$

millä tahansa $\boldsymbol{x} \in \mathbb{C}^n$.

Positiivisesti semidefiniittin matriisin tapaus osoitetaan todeksi samalla tavalla kuin positiivisesti definiitti tapaus, joten sivuutamme kyseisen todistuksen. □

Huomaamme seuraavassa lauseessa, että mielivaltaisesta matriisista tietyllä tavalla muodostetut matriisit ovat aina positiivisesti definiittejä. Tällä tulee olemaan tärkeä merkitys etenkin luvussa 5.1.

Lause 4.10. *Olkoon $A \in \mathbb{C}^{m\times n}$. Tällöin $A^*A \in \mathbb{C}^{n\times n}$ ja $AA^* \in \mathbb{C}^{m\times m}$ ovat positiivisesti semidefiniittejä matriiseja.*

Todistus. Lauseen 4.1 nojalla A^*A ja AA^* ovat hermiittisiä. Nyt

$$\langle A^*A\boldsymbol{x}, \boldsymbol{x}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle \geq 0$$

ja

$$\langle AA^*\boldsymbol{x}, \boldsymbol{x}\rangle = \langle A^*\boldsymbol{x}, A^*\boldsymbol{x}\rangle \geq 0,$$

joten lauseen 4.9 nojalla A^*A ja AA^* ovat positiivisesti semidefiniittejä. □

Huomataan, että lauseiden 4.8 ja 4.10 nojalla jokaista mielivaltaista matriisia $A \in \mathbb{C}^{m\times n}$ kohden on aina olemassa positiivisesti definiitti matriisi $\sqrt{A^*A} \in \mathbb{C}^{n\times n}$. Luvussa 5.1 tulemme käsittelemään näin muodostettuja matriiseja ja niiden ominaisarvoja.

Lause 4.11. *Jokaiselle matriisille $A \in \mathbb{C}^{m\times n}$ pätee*

$$\operatorname{Ker}(A^*A) = \operatorname{Ker} A \qquad \textit{ja} \qquad \operatorname{Ran}(A^*A) = \operatorname{Ran} A^*.$$

Todistus. Osoitetaan ensin, että $\operatorname{Ker}(A^*A) = \operatorname{Ker} A$. Olkoon $\boldsymbol{x} \in \mathbb{C}^n$. Oletetaan, että $\boldsymbol{x} \in \operatorname{Ker} A$, eli $A\boldsymbol{x} = \Theta$. Tällöin $\boldsymbol{x} \in \operatorname{Ker}(A^*A)$, koska $A^*A\boldsymbol{x} = A^*\Theta = \Theta$, joten $\operatorname{Ker} A \subseteq \operatorname{Ker}(A^*A)$. Oletetaan sitten, että $\boldsymbol{x} \in \operatorname{Ker} A^*A$, eli $A^*A\boldsymbol{x} = \Theta$. Tällöin kertomalla edellisen yhtälön molemmat puolet vasemmalta vektorilla $\boldsymbol{x}^*$ saamme $\boldsymbol{x}^*A^*A\boldsymbol{x} = \boldsymbol{x}^*\Theta = 0$, josta seuraa $\langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = 0$. Tämä yhtälö pätee vain, jos $A\boldsymbol{x} = \Theta$, joten $\boldsymbol{x} \in \operatorname{Ker} A$. Osoitimme siis, että $\operatorname{Ker}(A^*A) \subseteq \operatorname{Ker} A$, joten $\operatorname{Ker}(A^*A) = \operatorname{Ker} A$.

Todistetaan vielä, että $\operatorname{Ran}(A^*A) = \operatorname{Ran} A^*$. Jos $\boldsymbol{x} \in \operatorname{Ran}(A^*A)$, eli $A^*A\boldsymbol{y} = \boldsymbol{x}$, jollakin $\boldsymbol{y} \in \mathbb{C}^n$, niin selvästi myös $\boldsymbol{x} \in \operatorname{Ran} A^*$. Tällöin siis $\operatorname{Ran}(A^*A) \subseteq \operatorname{Ran} A^*$. Nyt vektoriavaruus $\operatorname{Ran}(A^*A)$ on vektoriavaruuden $\operatorname{Ran} A^*$ aliavaruus. Koska $\operatorname{Ker}(A^*A) = \operatorname{Ker} A$, niin

$$\operatorname{rank}(A^*A) = n - \dim\left[\operatorname{Ker}(A^*A)\right] = n - \dim\left[\operatorname{Ker}(A)\right] = \operatorname{rank} A.$$

Nyt lauseen 2.2 nojalla

$$\operatorname{rank}(A^*A) = \operatorname{rank} A = \operatorname{rank} A^*,$$

joten lauseen 2.3 nojalla on oltava $\operatorname{Ran}(A^*A) = \operatorname{Ran}(A^*)$. $\square$

5 Singulaariarvohajotelma

Tutkiessamme matriisien ominaisuuksia voimme usein tutkia niiden ominaisarvoja. Kuitenkin ominaisarvoja voidaan löytää vain neliömatriiseille, joten mielivaltaisen matriisin tutkiminen ei onnistu ominaisarvojen avulla. Emme myöskään voi tehdä mielivaltaiselle matriisille mielekästä hajotelmaa ominaisarvojen avulla. On kuitenkin mahdollista löytää mielivaltaiselle matriisille siitä kuvaavia skalaareita ja niiden avulla mielivaltaiselle matriisille mielekäs hajotelma. Tämä luku käsittelee näitä aiheita.

5.1 Singulaariarvot

Huomasimme edellisessä luvussa, että matriisi $\sqrt{A^*A} \in \mathbb{C}^{n\times n}$ voidaan aina muodostaa matriisista $A \in \mathbb{C}^{m\times n}$ ja lisäksi näin muodostettu matriisi on lisäksi aina positiivisesti semidefiniitti. Tässä luvussa käsittelemme kyseisen matriisin ominaisarvoja.

Määritelmä 5.1. Olkoot $A \in \mathbb{C}^{m\times n}$ ja $\sqrt{A^*A}$ siitä muodostettu positiivisesti semidefiniitti matriisi ja olkoot $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$ matriisin $\sqrt{A^*A}$ ominaisarvot. Tällöin matriisin $\sqrt{A^*A}$ ominaisarvoja $\lambda_1, \ldots, \lambda_r$ kutsutaan *matriisin A singulaariarvoiksi* ja merkitään $\lambda_i = s_i$, kun $i = 1, 2, \ldots, r$.

Koska matriisi $\sqrt{A^*A}$ on positiivisesti semidefiniitti, niin matriisin A singulaariarvot ovat aina positiivisia, eli $s_i > 0$, kun $i = 1, 2, \ldots, r$. Usein matriisin $\sqrt{A^*A}$ ominaisarvot $\lambda_{r+1} = \cdots = \lambda_n = 0$ määritellään myös matriisin A singulaariarvoiksi, mutta tämä määrittely osoittautuu usein ongelmalliseksi. Jos ottaisimme nollat mukaan singulaariarvoihin, joutuisimme kuitenkin lähes joka lauseessa käsittelemään vain nollasta eroavia singulaariarvoja.

Esimerkki 5.1. Etsitään matriisin

$$A = \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix}$$

singulaariarvot. Nyt

$$A^*A = \begin{bmatrix} -i & 0 & -i \\ 0 & -i & 0 \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix},$$

joten

$$\sqrt{A^*A} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{bmatrix}.$$

Koska $\sqrt{A^*A}$ on diagonaalimatriisi, on sen ominaisarvot sen diagonaalialkiot, eli $\lambda_1 = \sqrt{2}$ ja $\lambda_2 = 1$. Nyt siis matriisin A singulaariarvot ovat $s_1 = \sqrt{2}$ ja $s_2 = 1$.

Joskus singulaariarvot määritellään matriisin $\sqrt{AA^*}$ nollasta eroavina ominaisarvoina, mutta tällä ei ole merkittävää eroa määritelmäämme, kuten seuraava lause osoittaa.

Lause 5.1. *Olkoon $A \in \mathbb{C}^{m\times n}$. Tälloin matriisien $\sqrt{A^*A} \in \mathbb{C}^{n\times n}$ ja $\sqrt{AA^*} \in \mathbb{C}^{m\times m}$ nollasta eroavat ominaisarvot vastaavat toisiaan.*

Todistus. Lauseen todistamiseksi meidän tarvitsee todistaa lause matriiseille A^*A ja AA^*, jolloin saamme halutun lauseen todistettua myös neliöjuurimatriiseille. Olkoot $\lambda_1, \ldots, \lambda_n$ matriisin A^*A ominaisarvot ja olkoot $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ kyseisiä ominaisarvoja vastaavat ominaisvektorit siten, että ne muodostavat vektoriavaruuden $\mathbb{C}^n$ ortonormaalin kannan. Tällöin saamme

$$\langle A\boldsymbol{x_i}, A\boldsymbol{x_i}\rangle = \langle A^*A\boldsymbol{x_i}, \boldsymbol{x_i}\rangle = \langle \lambda_i\boldsymbol{x_i}, \boldsymbol{x_i}\rangle = \lambda_i \langle \boldsymbol{x_i}, \boldsymbol{x_i}\rangle = \lambda_i,$$

kun $i = 1, 2, \ldots, n$. Täten siis $A\boldsymbol{x_i} = \Theta$, jos ja vain jos $\lambda_i = 0$. Nyt

$$AA^*(A\boldsymbol{x_i}) = A(A^*A\boldsymbol{x_i}) = A(\lambda_i\boldsymbol{x_i}) = \lambda_i(A\boldsymbol{x_i}),$$

kun $i = 1, 2, \ldots, n$. Tällöin huomaamme, että jos $\lambda_i \neq 0$, niin $A\boldsymbol{x_i}$ on matriisin AA^* ominaisvektori. Jos siis $\lambda_i \neq 0$ on matriisin A^*A ominaisvektoria $\boldsymbol{x_i}$ vastaava ominaisarvo, niin λ_i on matriisin AA^* ominaisvektoria $A\boldsymbol{x_i}$ vastaava ominaisarvo. Tällöin siis matriisin A^*A ominaisarvojen joukko on matriisin AA^* ominaisarvojen joukon osajoukko. Osajoukkous toiseen suuntaan voidaan todistaa vaihtamalla edellisen todistuksen matriisien A ja A^* paikat keskenään. Tällöin matriisin A^*A nollasta eroavia ominaisarvoja vastaavat ominaisvektorit ovat muotoa $A^*\boldsymbol{x_i}$. Täten olemme siis todistaneet, että matriisien A^*A ja AA^* nollasta eroavat ominaisarvot vastaavat toisiaan, joten myös matriisien $\sqrt{A^*A}$ ja $\sqrt{AA^*}$ nollasta eroavat ominaisarvot vastaavat toisiaan. □

Nyt siis voisimme määritellä matriisin A singulaariarvot myös matriisin $\sqrt{AA^*}$ nollasta eroavina ominaisarvoina, kuten esimerkiksi kirjassa [1] tehdään. Tämä myös helpottaa joissakin tapauksissa singulaariarvojen laskemista.

Lause 5.2. *Olkoon $A \in \mathbb{C}^{m\times n}$ ja olkoot $\boldsymbol{x_1}, \ldots, \boldsymbol{x_r}$ matriisin A^*A (vaihtoehtoisesti AA^*) ortonormaalit ominaisvektorit, jotka vastaavat sen nollasta eroavia ominaisarvoja. Tällöin $A\boldsymbol{x_1}, \ldots, A\boldsymbol{x_r}$ ($A^*\boldsymbol{x_1}, \ldots, A^*\boldsymbol{x_r}$) ovat matriisin AA^* (A^*A) nollasta erovia ominaisarvoja vastaavat ortogonaaliset ominaisvektorit.*

Todistus. Lauseessa 5.1 osoitimme jo, että matriisien A^*A ja AA^* nollasta eroavat ominaisarvot vastaavat toisiaan. Samassa lauseessa osoitimme myös, että jos $\boldsymbol{x_1}$ on matriisin A^*A nollasta eroavaa ominaisarvoa vastaava ominaisvektori, niin $A\boldsymbol{x_i}$ on matriisin AA^* samaa ominaisarvoa vastaava ominaisvektori. Olkoot $\boldsymbol{x_1}, \ldots, \boldsymbol{x_r}$ matriisin A^*A ortonormaalit ominaisvektorit, jotka vastaavat sen nollasta eroavia ominaisarvoja, ja olkoot $\boldsymbol{y_1}, \ldots, \boldsymbol{y_r}$ matriisin AA^* samoja nollasta eroavia ominaisarvoja vastaavat ominaisvektorit, jolloin siis $\boldsymbol{y_i} = A\boldsymbol{x_i}$, kun $i = 1, 2, \ldots, r$. Nyt

$$\langle \boldsymbol{y_i}, \boldsymbol{y_j} \rangle = \langle A\boldsymbol{x_i}, A\boldsymbol{x_j} \rangle = \langle A^*A\boldsymbol{x_i}, \boldsymbol{x_j} \rangle = \langle \lambda_i \boldsymbol{x_i}, \boldsymbol{x_j} \rangle = \lambda_i \langle \boldsymbol{x_i}, \boldsymbol{x_j} \rangle = 0,$$

kun $i \neq j$, $i = 1, 2, \ldots, r$ ja $j = 1, 2, \ldots, r$. Täten siis vektorit $\boldsymbol{y_1}, \ldots, \boldsymbol{y_r}$ ovat ortogonaalisia. Vastaava väite saadaan vaihtamalle matriisien A ja A^* paikkoja, joten sivuutamme kyseisen todistuksen. □

Jos siis tiedämme matriisin A^*A ortonormaalit ominaisvektorit, jotka ovat aina olemassa, koska A^*A on normaali, voimme ratkaista matriisin AA^* nollasta erovia ominaisarvoja vastaavat ortonormaalit ominaisvektorit kertomalla matriisin A^*A ortonormaalit ominaisvektorit matriisilla A ja jakamalla ne normillansa. Tämä matriisijoukko voidaan lisäksi täydentää matriisin AA^* ortonormaaliksi ominaisvektorikannaksi lisäämällä siihen matriisin AA^* ytimen ortonormaali kanta.

5.2 Singulaariarvohajotelma

Tässä kappaleessa käsittelemme singulaariarvohajotelmaa, joka liittyy vahvasti singulaariarvoihin. Singulaariarvohajotelma on erittäin mielekäs työkalu mielivaltaisten matriisien tutkimiseen, ja sille löytyykin paljon sovelluksia esimerkiksi tilastotieteessä.

Lause 5.3 (Singulaariarvohajotelma)**.** *Olkoon $A \in \mathbb{C}^{m \times n}$ ja olkoot $s_1, \ldots, s_r$ matriisin A singulaariarvot. Tällöin matriisi A voidaan esittää muodossa*

$$A = UDV^*,$$

jossa $U \in \mathbb{C}^{m \times m}$ ja $V \in \mathbb{C}^{n \times n}$ ovat unitaarisia ja $D \in \mathbb{C}^{m \times n}$ on matriisi, jossa $D(i,i) = s_i$, kun $i = 1, 2, \ldots, r$, ja sen muut alkiot ovat nollia.

Todistus. Olkoot $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$ matriisin A^*A ominaisarvot ja olkoot $\boldsymbol{x_1}, \ldots \boldsymbol{x_n}$ niitä vastaavat ominaisvektorit, jotka muodostavat vektoriavaruuden $\mathbb{C}^n$ ortonormaalin kannan.

Koska lauseen 5.2 nojalla matriisin AA^* nollasta eroavia ominaisarvoja vastaavat ominaisvektorit ovat ortogonaalisia, niin saamme matriisin AA^* ominaisarvoja $\lambda_1, \ldots, \lambda_r$ vastaavat ortonormaalit ominaisvektorit kaavalla

$$\begin{aligned}\boldsymbol{y_i} &= \frac{1}{\|A\boldsymbol{x_i}\|} A\boldsymbol{x_i} = \frac{1}{\sqrt{\langle A\boldsymbol{x_i}, A\boldsymbol{x_i} \rangle}} A\boldsymbol{x_i} = \frac{1}{\sqrt{\langle A^*A\boldsymbol{x_i}, \boldsymbol{x_i} \rangle}} A\boldsymbol{x_i} \\ &= \frac{1}{\sqrt{\langle \lambda_i \boldsymbol{x_i}, \boldsymbol{x_i} \rangle}} A\boldsymbol{x_i} = \frac{1}{\sqrt{\lambda_i}\sqrt{\langle \boldsymbol{x_i}, \boldsymbol{x_i} \rangle}} A\boldsymbol{x_i} = \frac{1}{\sqrt{\lambda_i}} A\boldsymbol{x_i},\end{aligned}$$

kun $i = 1, 2, \ldots, r$. Valitaan matriisin AA^* ytimestä vektorit $\boldsymbol{y_{r+1}}, \ldots, \boldsymbol{y_m}$, jotka muodostavat ytimen ortonormaalin kannan, jolloin vektorit $\boldsymbol{y_1}, \ldots, \boldsymbol{y_m}$ muodostavat matriisin AA^* ortonormaalin ominaisvektorikannan. Nyt saamme edellisestä yhtälöstä

$$A\boldsymbol{x_i} = \sqrt{\lambda_i}\boldsymbol{y_i} = s_i\boldsymbol{y_i}, \tag{5.1}$$

kun $i = 1, 2, \ldots, r$, ja määritelmänsä mukaan $A\boldsymbol{x_i} = \Theta$, kun $i = r + 1, \ldots, n$.

Olkoot nyt matriisit V ja U muotoa

$$V = \begin{bmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \cdots & \boldsymbol{x_n} \end{bmatrix}$$

ja

$$U = \begin{bmatrix} \boldsymbol{y_1} & \boldsymbol{y_2} & \cdots & \boldsymbol{y_m} \end{bmatrix},$$

jolloin lauseen 3.1 nojalla ne ovat unitaarisia. Nyt yhtälön (5.1) avulla voimme muodostaa yhtälön

$$AV = \begin{bmatrix} A\boldsymbol{x_1} & \cdots & A\boldsymbol{x_n} \end{bmatrix} = \begin{bmatrix} s_1\boldsymbol{y_1} & \cdots & s_r\boldsymbol{y_r} & \Theta & \cdots & \Theta \end{bmatrix} = UD,$$

josta saamme halutun muodon $A = UDV^*$. □

Singulaariarvohajotelman todistuksen perusteella voimme sanoa enemmän matriiseista U ja V. Matriisi U muodostuu matriisin AA^* ortonormaaleista ominaisvektoreista ja matriisi V matriisin A^*A ortonormaaleista ominaisvektoreista.

Määritelmä 5.2. Lauseen 5.3 todistuksessa esiintyviä ortonormaaleja vektoreita $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$ ja $\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_m}$ kutsutaan *matriisin A oikean- ja vasemmanpuoleisiksi singulaarivektoreiksi.* Nämä vektorit muodostavat vektoriavaruuksien $\mathbb{C}^n$ ja $\mathbb{C}^m$ ortonormaalit kannat, joita kutsutaan *matriisin A oikean- ja vasemmanpuoleisiksi singulaarikannoiksi.*

Määritelmässä esiintyvät vektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$ siis muodostavat matriisin A^*A ortonormaalin ominaisvektorikannan ja vektorit $\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_m}$ muodostavat matriisin AA^* ortonormaalin ominaisvektorikannan, eli matriisien V ja U sarakkeet.

Seuraava lause kokoaa myös hieman lauseen 5.3 sisältöä ja antaa hieman lisää laskennallisia työkaluja singulaariarvohajotelmien ratkaisemiseen.

Lause 5.4. *Muodostakoon vektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$ ja $\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_m}$ matriisin $A \in \mathbb{C}^{m\times n}$ singulaarikannat ja olkoot $s_1, s_2, \ldots, s_r$ matriisin A singulaariarvot. Tällöin*

$$A\boldsymbol{x_i} = \begin{cases} s_i\boldsymbol{y_i}, & \text{kun } i = 1, 2, \ldots, r \\ \Theta, & \text{kun } i = r + 1, \ldots, n \end{cases}$$

ja

$$A^*\boldsymbol{y_i} = \begin{cases} s_i\boldsymbol{x_i}, & \text{kun } i = 1, 2, \ldots, r \\ \Theta, & \text{kun } i = r + 1, \ldots, m. \end{cases}$$

Todistus. Lauseen ensimmäinen yhtälö osoitettiin todeksi jo lauseen 5.3 todistuksen yhtälössä (5.1), joten osoitetaan vielä, että toinenkin yhtälö on tosi. Koska vektorit $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_r}$ ovat matriisin A^*A ominaisvektoreita, niille pätee $A^*A\boldsymbol{x_i} = \lambda_i \boldsymbol{x_i} = s_i^2 \boldsymbol{x_i}$, kun $i = 1, 2, \dots, n$. Tällöin kertomalla ensimmäinen yhtälön molemmat puolet vasemmalta matriisilla A^* ja jakamalla yhtälön molemmat puolet singulaariarvolla s_i saamme halutun yhtälön

$$A^*\boldsymbol{y_i} = \begin{cases} s_i \boldsymbol{x_i}, & \text{kun } i = 1, 2, \dots, r \\ \Theta, & \text{kun } i = r+1, \dots, m. \end{cases}$$

□

Vaikka jokaisella mielivaltaisella matriisilla on olemassa jokin singulaariarvohajotelma, niin singulaariarvohajotelma ei ole välttämättä yksikäsitteinen, kuten seuraavassa esimerkissä huomaamme.

Esimerkki 5.2. Tarkastellaan vielä esimerkin 5.1 matriisia

$$A = \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix},$$

jonka singulaariarvot ovat $s_1 = \sqrt{2}$ ja $s_2 = 1$. Nyt matriisi A^*A on diagonaalimatriisi, joten voimme valita sen ominaisvektorikannaksi vektoriavaruuden $\mathbb{C}^2$ luonnollisen kannan, jolloin $V = I$. Nyt lauseen 5.4 avulla saamme

$$\boldsymbol{y_1} = \frac{A\boldsymbol{x_1}}{s_1} = \frac{\begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ i \end{bmatrix}$$

ja

$$\boldsymbol{y_2} = \frac{A\boldsymbol{x_2}}{s_2} = \frac{\begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{1} = \begin{bmatrix} 0 \\ i \\ 0 \end{bmatrix}.$$

Vektorin $\boldsymbol{y_3}$ tulee olla ortonormaali vektoreiden $\boldsymbol{y_1}$ ja $\boldsymbol{y_2}$ kanssa. Täten voimme valita matriisin AA^* ytimestä vektorin

$$\boldsymbol{y_3} = \frac{1}{\sqrt{2}} \begin{bmatrix} i \\ 0 \\ -i \end{bmatrix},$$

jolloin

$$U_1 = \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ 0 & i & 0 \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{bmatrix}.$$

Nyt siis matriisin A singulaariarvohajotelma on

$$U_1DV^* = \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ 0 & i & 0 \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} = A.$$

Toisaalta voimme valita myös matriisin AA^* ytimestä vektorin

$$\boldsymbol{y_4} = -\boldsymbol{y_3} = \frac{1}{\sqrt{2}} \begin{bmatrix} -i \\ 0 \\ i \end{bmatrix},$$

jolloin

$$U_2 = \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & i & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \end{bmatrix}.$$

Tällöin matriisin A singulaariarvohajotelma on

$$U_2DV^* = \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & i & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} = A.$$

Selvästi siis matriisin A singulaariarvohajotelma ei ole yksikäsitteinen.

6 Pseudoinverssi ja pienimmän neliösumman probleema

Tässä luvussa käsittelemme singulaariarvohajotelman avulla mielivaltaisen matriisin yleistettyä käänteismatriisia eli pseudoinverssiä (pseudoinverssi voidaan käsitellä myös ilman tietoa singulaariarvohajotelmasta kuten kirjassa [3]). Lisäksi käsittelemme pseudoinverssin käyttöä pienimmän neliösumman probleeman ratkaisussa ja löydämme mille tahansa lineaariselle yhtälöryhmälle pienimmän neliösumman parhaan ratkaisuvektorin.

6.1 Pseudoinverssi

Neliömatriisilla A, jolle $\det A \neq 0$, on aina olemassa käänteismatriisi A^{-1}, jolle $A^{-1}A = AA^{-1} = I$. Jos $\det A = 0$ tai matriisi A ei ole neliömatriisi, niin tällaista matriisia ei ole mahdollista löytää. Esiin nousee kysymys, onko mielivaltaiselle matriisille olemassa jonkinlaista käänteismatriisia, jolle olisi samantyyppisiä ominaisuuksia kuin neliömatriisin A käänteismatriisilla A^{-1}. Voimme löytää tällaisen matriisin, tosin ei ihan samanlaisilla ominaisuuksilla. Kehitämme seuraavaksi tällaisen yleisen käänteismatriisin. Lisää teoriaa käänteismatriiseista löytyy esimerkiksi kirjasta [3, s. 424–432].

Määritelmä 6.1. Olkoon $A \in \mathbb{C}^{m\times n}$. Tällöin ehdot

(1) $(AX)^* = AX$, $(XA)^* = XA$,

(2) $AXA = A$ ja

(3) $XAX = X$

toteuttavaa matriisia $X \in \mathbb{C}^{n\times m}$ kutsutaan *matriisin A Moore–Penrosen yleistetyksi käänteismatriisiksi* tai lyhyemmin *matriisin A pseudoinversiksi* ja merkitään $X = A^+$.

Heti ensimmäisenä esiin nousee kysymys, onko pseudoinverssi olemassa mille tahansa mielivaltaiselle matriisille? Seuraavassa lauseessa osoitamme, että pseudoinverssi todellakin löytyy mille tahansa matriisille.

Lause 6.1. *Olkoot alkiot $s_1, \ldots, s_r$ matriisin $A \in \mathbb{C}^{m\times n}$ singulaariarvot ja olkoon $A = UDV^*$ matriisin A singulaariarvohajotelma. Tällöin matriisi $A^+ = VD^+U^*$, jonka matriisi $D^+ \in \mathbb{C}^{n\times m}$ on matriisin D transpoosi, jossa on alkioiden $s_1, \ldots, s_r$ paikalle vaihdettu niiden käänteisalkiot $s_1^{-1}, \ldots, s_r^{-1}$, on matriisin A pseudoinverssi.*

Todistus. Tarkastellaan ensin matriiseja $D^+D \in \mathbb{C}^{n\times n}$ ja $DD^+ \in \mathbb{C}^{m\times m}$. Ne ovat diagonaalimatriiseja, joissa $D^+D(i,i) = DD^+(i,i) = 1$, kun $i = 1, 2, \ldots, r$, ja muut alkiot ovat nollia. Tällöin selvästi D^+D ja DD^+ ovat hermiittisiä, $DD^+D = D$ ja $D^+DD^+ = D^+$, joten D^+ on matriisin D pseudoinverssi. Nyt

$$\begin{aligned}(AA^+)^* &= (UDV^*VD^+U^*)^* = (UDD^+U^*)^* \\ &= U(DD^+)^*U^* = UDD^+U^* = UDV^*VD^+U^* = AA^+\end{aligned}$$

ja

$$\begin{aligned}(A^+A)^* &= (VD^+U^*UDV^*)^* = (VD^+DV^*)^* \\ &= V(D^+D)^*V^* = V(D^+D)V^* = VD^+U^*UDV^* = A^+A,\end{aligned}$$

joten ehto (1) on voimassa. Seuraavaksi

$$AA^+A = UDV^*VD^+U^*UDV^* = UDD^+DV^* = UDV^* = A,$$

joten myös ehto (2) on voimassa. Lopuksi

$$A^+AA^+ = VD^+U^*UDV^*VD^+U^* = UD^+DD^+V^* = UD^+V^* = A^+,$$

joten ehto (3) on voimassa ja lause on todistettu. □

Edellä osoitimme, että matriisille on aina olemassa pseudoinverssi. Voisimmeko mahdollisesti löytää toisenlaisen pseudoinverssin samalle matriisille? Tämä ei ole mahdollista, sillä matriisin pseudoinverssi on aina yksikäsitteinen. Osoitamme tämän seuraavassa lauseessa.

Lause 6.2. *Määritelmän 6.1 ehdot (1)–(3) toteuttava matriisi on yksikäsitteinen.*

Todistus. Olkoot matriisit X_1 ja X_2 määritelmän 6.1 ehdot (1)–(3) toteuttavat matriisit. Tällöin ehtojen (1) ja (3) avulla saamme

$$X_i^* = X_i^*A^*X_i^* = (AX_i)^*X_i^* = AX_iX_i^*,$$

kun $i = 1, 2$. Nyt saamme

$$X_1^* - X_2^* = A(X_1X_1^* - X_2X_2^*),$$

joten $\mathrm{Ran}\,(X_1^* - X_2^*) \subseteq \mathrm{Ran}\,A$.

Toisaalta ehtojen (1) ja (2) avulla saamme

$$A^* = A^*X_i^*A^* = (X_iA)^*A^* = X_iAA^*,$$

kun $i = 1, 2$. Nyt $X_iAA^* - A^* = (X_iA - I)A^* = \mathbf{0}$, josta seuraa myös, että $A(A^*X_i^* - I) = \mathbf{0}$. Täten saamme

$$A(A^*X_1^* - I) - A(A^*X_2^* - I) = A(A^*X_1^* - I - A^*X_2^* + I) = AA^*(X_1^* - X_2^*) = \mathbf{0},$$

joten $\mathrm{Ran}\,(X_1^* - X_2^*) \subseteq \mathrm{Ker}\,AA^*$. Nyt lauseen 4.11 nojalla $\mathrm{Ker}\,AA^* = \mathrm{Ker}\,A^*$, joten $\mathrm{Ran}\,(X_1^* - X_2^*) \subseteq \mathrm{Ker}\,A^*$. Koska $\mathrm{Ran}\,A$ ja $\mathrm{Ker}\,A^*$ ovat toistensa ortogonaalikomplementteja lauseen 2.15 nojalla, niin tulee olla $X_1^* - X_2^* = \mathbf{0}$, eli $X_1^* = X_2^*$. Tällöin siis matriisi X on yksikäsitteinen. □

Vaikka siis osoitimme esimerkissä 5.2, ettei mielivaltaisen matriisin singulaariarvohajotelma ole välttämättä yksikäsitteinen, niin silti sen pseudoinverssi on lauseiden 6.1 ja 6.2 nojalla aina yksikäsitteinen. Seuraava esimerkki osoittaa tämän esimerkin 5.2 matriisin tapauksessa.

Esimerkki 6.1. Esimerkissä 5.2 saimme matriisin A singulaariarvohajotelmiksi sekä

$$U_1DV^* = \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ 0 & i & 0 \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} = A$$

että

$$U_2DV^* = \begin{bmatrix} \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & i & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & i \\ i & 0 \end{bmatrix} = A.$$

Nyt ensimmäisen singulaariarvohajotelman avulla saadaan

$$A^+ = VD^+U_1^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & -i & 0 \\ -\frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{i}{2} & 0 & -\frac{i}{2} \\ 0 & -i & 0 \end{bmatrix}.$$

Toisaalta toisen singulaariarvohajotelman avulla saadaan myös

$$A^+ = VD^+U_2^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -\frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & -i & 0 \\ \frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{i}{2} & 0 & -\frac{i}{2} \\ 0 & -i & 0 \end{bmatrix}.$$

Ratkaistaksemme matriisin pseudoinverssin meidän ei aina tarvitse tietää matriisin singulaariarvohajotelmaa. Jos matriisilla on käänteismatriisi, niin se on myös kyseisen matriisin pseudoinverssi, sillä selvästi $AA^{-1}A = A$, $A^{-1}AA^{-1} = A^{-1}$ ja $(AA^{-1})^* = (A^{-1}A)^* = I = A^{-1}A = AA^{-1}$. Myös muita erikoistapauksia on olemassa, ja käsittelemme niitä seuraavassa lauseessa.

Lause 6.3. *Olkoon* $A \in \mathbb{C}^{m\times n}$ *ja* $\operatorname{rank} A = \min(m,n)$*. Jos* $m \le n$ *ja* AA^* *on kääntyvä, niin* $A^+ = A^*(AA^*)^{-1}$*. Jos taas* $n \le m$ *ja* A^*A *on kääntyvä, niin* $A^+ = (A^*A)^{-1}A^*$*.*

Todistus. Oletetaan, että $m \le n$ ja $\operatorname{rank} A = m$. Tällöin $\operatorname{rank} A^* = m$ ja $\operatorname{rank} AA^* = \min(\operatorname{rank} A, \operatorname{rank} A^*) = m$. Koska $AA^* \in \mathbb{C}^{m\times m}$, niin $(AA^*)^{-1}$ on olemassa ja myös matriisi $A^*(AA^*)^{-1}$ on olemassa. Nyt saamme selvästi $(AA^*(AA^*)^{-1})^* = I^* = I$ ja

$$(A^*(AA^*)^{-1}A)^* = A^*((AA^*)^{-1})^*A = A^*((AA^*)^*)^{-1}A = A^*(AA^*)^{-1}A,$$

joten ehto (1) on voimassa. Huomaamme lisäksi, että $AA^*(AA^*)^{-1}A = A$ ja $A^*(AA^*)^{-1}AA^*(AA^*)^{-1} = A(AA^*)^{-1}$, joten ehdot (2) ja (3) ovat voimassa. Täten ensimmäinen väite on tosi.

Toinen väite saadaan lähes identtisillä vaiheilla kuin ensimmäinen, joten sivuutamme kyseisen todistuksen. □

Seuraavassa lauseessa esitämme joitakin hyödyllisiä pseudoinverssien ominaisuuksia.

Lause 6.4. *Olkoon matriisi $A^+ \in \mathbb{C}^{n\times m}$ matriisin $A \in \mathbb{C}^{m\times n}$ pseudoinverssi. Tällöin*

(1) $(A^*)^+ = (A^+)^*$*,*

(2) $(aA)^+ = a^{-1}A^+$*, kun* $a \in \mathbb{C}$ *ja* $a \neq 0$*,*

(3) jos A on normaali (hermiittinen, positiivisesti semidefiniitti), niin samoin on myös A^+,

(4) $(UAV)^+ = V^*A^+U^*$*, jos* $U \in \mathbb{C}^{m\times m}$ *ja* $V \in \mathbb{C}^{n\times n}$ *ovat unitaarisia matriiseja,*

(5) $(A^+)^+ = A$*,*

(6) $(A^k)^+ = (A^+)^k$*, kun* $k \in \mathbb{Z}_+$ *ja A on normaali matriisi,*

(7) $(AA^*)^+ = (A^+)^*A^+$ *ja* $(A^*A)^+ = A^+(A^+)^*$*.*

Todistus. Todistetaan ensin kohta (2) pseudoinverssin määritelmän avulla. Nyt $a^{-1}A^+aA = A^+A$ ja $aAa^{-1}A^+ = AA^+$, joten selvästi $a^{-1}A^+aA$ ja $aAa^{-1}A^+$ ovat hermiittisiä. Koska lisäksi

$$a^{-1}A^+aAa^{-1}A^+ = a^{-1}A^+AA^+ = a^{-1}A^+$$

ja

$$aAa^{-1}A^+aA = aAA^+A = aA,$$

niin $(aA)^+ = a^{-1}A^+$.

Todistetaan seuraavaksi kohta (4) pseudoinverssin määritelmän avulla, jotta voimme käyttää tulosta muiden kohtien todistamiseen. Koska

$$V^*A^+U^*UAV = (V^*(A^+A)^*V)^* = (V^*A^+AV)^* = (V^*A^+U^*UAV)^*$$

ja

$$UAVV^*A^+U^* = (U(AA^+)^*U^*)^* = (UAA^+U^*)^* = (UAVV^*A^+U^*)^*,$$

niin määritelmän kohta (1) on voimassa. Nyt lisäksi

$$V^*A^+U^*UAVV^*A^+U^* = V^*A^+AA^+U^* = V^*A^+U^*$$

ja

$$UAVV^*A^+U^*UAV = UAA^+AV = UAV,$$

joten määritelmän kohdat (2) ja (3) ovat voimassa. Täten unitaaristen matriisien U ja V ollessa sopivan kokoisia $(UAV)^+ = V^*A^+U^*$.

Todistetaan sitten kohta (1) kohdan (4) tuloksen avulla. Olkoon $A = UDV^*$ matriisin A singulaariarvohajotelma, jolloin $A^+ = VD^+U^*$. Nyt

$$(A^+)^* = (VD^+U^*)^* = U(D^+)^*V^* = U(D^*)^+V^* = (VD^*U^*)^+ = (A^*)^+,$$

koska $(D^+)^* = (D^*)^+$, sillä molemmissa tapauksissa matriisia D on transponoitu kaksi kertaa ja lisäksi vaihdettu diagonaalille alkioiden $s_1, \dots, s_r$ paikoille alkiot $s_1^{-1}, \dots, s_r^{-1}$.

Osoitetaan seuraavaksi kohta (3) todeksi. Oletetaan, että matriisi $A \in \mathbb{C}^{n\times n}$ on normaali. Tällöin se voidaan esittää muodossa $A = UD_1U^*$, jossa $D_1 \in \mathbb{C}^{n\times n}$ on diagonaalimatriisi, jonka alkiot ovat matriisin A ominaisarvot, ja $U \in \mathbb{C}^{n\times n}$ on unitaarinen. Huomataan, että matriisin D_1 pseudoinverssi D_1^+ on sellainen matriisi, jossa matriisi D_1 on transponoitu ja alkiot λ_i on vaihdettu niiden käänteisalkioihin λ_i^{-1}, kun $1, 2, \dots, n$, paitsi jos $\lambda_i = 0$, jolloin kyseinen ominaisarvo pysyy nollana. Nyt kohdan (4) avulla saamme $A^+ = (UD_1U^*)^+ = UD_1^+U^*$. Tällöin siis A^+ on unitaarisesti diagonalisoituva, joten se on myös normaali. Jos A on hermiittinen, niin kohdan (1) nojalla $A^+ = (A^*)^+ = (A^+)^*$, joten tällöin myös A^+ on hermiittinen. Jos taas A on positiivisesti semidefiniitti, niin se on myös normaali. Tällöin saadaan taas $A = UD_1U^*$, josta seuraa, että $A^+ = UD_1^+U^*$. Koska matriisin A ominaisarvot ovat ei-negatiivisia, niin myös matriisin A^+ ominaisarvot ovat ei-negatiivisia, johtuen matriisin D_1^+ muodostustavasta, jossa positiivisista ominaisarvoista otetaan niiden käänteisluku ja nollat pysyvät nollina. Täten myös A^+ on positiivisesti semidefiniitti.

Kohta (5) on helppo osoittaa todeksi kohdan (4) avulla. Olkoon $A = UDV^*$ matriisin A singulaariarvohajotelma, jolloin $A^+ = VD^+U^*$. Tällöin $(A^+)^+ = (VD^+U^*)^+ = U(D^+)^+V^*$. Nyt selvästi $(D^+)^+ = D$, joten $(A^+)^+ = UDV^* = A$.

Tarkastellaan nyt kohtaa (6). Oletetaan, että A on normaali, jolloin voidaan taas kirjoittaa $A = UD_1U^*$. Tällöin

$$\begin{aligned}(A^k)^+ &= (\underbrace{A \cdots A}_{\text{k kpl}})^+ = (\underbrace{(UD_1U^*) \cdots (UD_1U^*)}_{\text{k kpl}})^+ \\ &= (U\underbrace{D_1 \cdots D_1}_{\text{k kpl}}U^*)^+ = (UD_1^kU^*)^+ = U(D_1^k)^+U^*.\end{aligned}$$

Nyt $D_1^k = \operatorname{diag}(\lambda_i^k)$, kun $i = 1, 2, \dots, n$, jolloin $(D_1^k)^+$ on sellainen matriisi, jossa matriisi D on transponoitu ja jokaisen diagonaalialkion paikalle on vaihdettu sen käänteisalkio λ_i^{-k}, paitsi jos $\lambda_i = 0$, jolloin kyseinen diagonaalialkio pysyy nollana. Tällöin selvästi $(D_1^k)^+ = (D_1^+)^k$, joten saamme

$$(A^k)^+ = U(D_1^k)^+U^* = U(D_1^+)^kU^* = \underbrace{(UD_1^+U^*) \cdots (UD_1^+U^*)}_{\text{k kpl}} = (A^+)^k.$$

Todistetaan vielä kohta (7), jonka jälkeen olemme todistaneet koko lauseen. Olkoon $A = UDV^*$ matriisin A singulaariarvohajotelma, jolloin $A^+ =$

VD^+U^*. Osoitetaan ensin, että $(D^*D)^+ = D^+(D^*)^+$ ja $(DD^*)^+ = (D^*)^+D^+$. Muistetaan, että $D(i,i) = s_i > 0$, kun $i = 1, 2, \dots, r$, ja muut matriisin D alkiot ovat nollia. Tällöin $D^*D \in \mathbb{C}^{n\times n}$ on matriisi, jossa $D^*D(i,i) = s_i^2$, kun $i = 1, 2, \dots, r$, ja muut alkiot ovat nollia. Nyt $(D^*D)^+$ on sellainen matriisi, jossa alkiot s_i^2 on muutettu alkioiksi s_i^{-2}, kun $i = 1, 2, \dots, r$, ja muut alkiot ovat nollia. Matriisissa $D^+(D^*)^+ \in \mathbb{C}^{n\times n}$ on diagonaalilla alkiot $s_i^{-1} \cdot s_i^{-1} = s_i^{-2}$, kun $i = 1, 2, \dots, r$, ja muut alkiot ovat nollia. Tällöin $(D^*D)^+ = D^+(D^*)^+$. Lähes samoilla argumenteilla voimme osoittaa, että $(DD^*)^+ = (D^*)^+D^+$, joten sivuutamme kyseisen todistuksen. Nyt

$$\begin{aligned}(AA^*)^+ &= (UDV^*VD^*U^*)^+ = U(DD^*)^+U^* \\ &= U(D^*)^+V^*VD^+U^* = (VD^+U^*)^*(VD^+U^*) = (A^+)^*A^+\end{aligned}$$

ja

$$\begin{aligned}(A^*A)^+ &= (VD^*U^*UDV^*)^+ = V(D^*D)^+V^* \\ &= VD^+U^*U(D^*)^+V^* = (VD^+U^*)(VD^+U^*)^* = A^+(A^+)^*,\end{aligned}$$

joten olemme todistaneet lauseen. □

Seuraavassa lauseessa käsitellään vielä yksi pseudoinversseihin liittyvä ominaisuus.

Lause 6.5. *Olkoon $A \in \mathbb{C}^{m\times n}$ ja olkoon $A^+ \in \mathbb{C}^{n\times m}$ matriisin A pseudoinverssi. Tällöin*

$$A^+A\boldsymbol{x} = \boldsymbol{x} \qquad \text{ja} \qquad AA^+\boldsymbol{y} = \boldsymbol{y}$$

kaikilla vektoreilla $\boldsymbol{x} \in \operatorname{Ran} A^*$ ja $\boldsymbol{y} \in \operatorname{Ran} A$.

Todistus. Oletetaan ensin, että $\boldsymbol{x} \in \operatorname{Ran} A^*$, jolloin $A^*\boldsymbol{z_1} = \boldsymbol{x}$, jollakin vektorilla $\boldsymbol{z_1} \in \mathbb{C}^m$. Tällöin määritelmän 6.1 ja lauseen 6.4 nojalla

$$A^+A\boldsymbol{x} = (A^+A)^*A^*\boldsymbol{z_1} = A^*(A^*)^+A^*\boldsymbol{z_1} = A^*\boldsymbol{z_1} = \boldsymbol{x},$$

joten ensimmäinen yhtälö on tosi.

Jos $\boldsymbol{y} \in \operatorname{Ran} A$, niin on olemassa vektori $\boldsymbol{z_2} \in \mathbb{C}^n$ siten, että $A\boldsymbol{z_2} = \boldsymbol{y}$. Nyt määritelmän 6.1 nojalla saamme

$$AA^+\boldsymbol{y} = AA^+A\boldsymbol{z_2} = A\boldsymbol{z_2} = \boldsymbol{y},$$

joten lause on tosi. □

Seuraava lause on jatkoa edellisessä luvussa käsitellystä lauseeseen 5.4 ja se nojautuu vahvasti pseudoinverssin esitysmuotoon $A = VD^+U^*$.

Lause 6.6. *Muodostakoon vektorit* $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}$ *ja* $\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_m}$ *matriisin* $A \in \mathbb{C}^{m \times n}$ *singulaarikannat ja olkoot* $s_1, s_2, \dots, s_r$ *matriisin* A *nollasta eroavat singulaariarvot. Tällöin*

$$A^+\boldsymbol{y_i} = \begin{cases} s_i^{-1}\boldsymbol{x_i}, & \textit{kun } i = 1, 2, \dots, r \\ \Theta, & \textit{kun } i = r+1, \dots, n \end{cases}$$

ja

$$(A^+)^*\boldsymbol{x_i} = \begin{cases} s_i^{-1}\boldsymbol{y_i}, & \textit{kun } i = 1, 2, \dots, r \\ \Theta, & \textit{kun } i = r+1, \dots, m. \end{cases}$$

Todistus. Aluksi huomataan, että lauseen 5.4 nojalla $\boldsymbol{x_i} \in \operatorname{Ran} A^*$ ja $\boldsymbol{y_i} \in \operatorname{Ran} A$, kun $i = 1, 2, \dots, r$. Tällöin lauseen 6.5 nojalla $\boldsymbol{x_i} = A^+A\boldsymbol{x_i}$. Nyt kertomalla lauseen 5.4 ensimmäinen yhtälön molemmat puolet vasemmalta matriisilla A^+ saamme

$$A^+A\boldsymbol{x_i} = \boldsymbol{x_i} = \begin{cases} s_iA^+\boldsymbol{y_i}, & \text{kun } i = 1, 2, \dots, r \\ \Theta, & \text{kun } i = r+1, \dots, m, \end{cases}$$

josta saamme halutun tuloksen kertomalla yhtälön luvulla s_i^{-1}.

Lauseesta 6.5 saamme

$$\boldsymbol{z} = AA^+\boldsymbol{z} = (AA^+)^*\boldsymbol{z} = (A^+)^*A^*\boldsymbol{z},$$

kun $\boldsymbol{z} \in \operatorname{Ran} A$. Koska $\boldsymbol{y_i} \in \operatorname{Ran} A$, niin kertomalla lauseen 5.4 toisen yhtälön molemmat puolet vasemmalta matriisilla $(A^+)^*$ saamme lauseen 6.5 nojalla

$$(A^+)^*A^*\boldsymbol{y_i} = \boldsymbol{y_i} = \begin{cases} s_i(A^+)^*\boldsymbol{x_i}, & \text{kun } i = 1, 2, \dots, r \\ \Theta, & \text{kun } i = r+1, \dots, n, \end{cases}$$

josta saamme toisen halutun tuloksen kertomalla yhtälön luvulla s_i^{-1}. □

Seuraavat kaksi lausetta käsittelevät pseudoinverssin ydintä ja kuvaa. Lopulta huomaamme, että niistä löytyy mielenkiintoinen yhteys alkuperäisen matriisin ja sen transpoosikonjugaatin ytimeen ja kuvaa.

Lause 6.7. *Olkoon* $A \in \mathbb{C}^{m \times n}$ *ja olkoon* $A^+ \in \mathbb{C}^{n \times m}$ *matriisin* A *pseudoinverssi. Tällöin*

$$\operatorname{Ker}(AA^+) = \operatorname{Ker} A^+ \qquad \textit{ja} \qquad \operatorname{Ran}(AA^+) = \operatorname{Ran} A.$$

Todistus. Osoitetaan ensin, että $\operatorname{Ker}(AA^+) = \operatorname{Ker} A^+$. Jos $\boldsymbol{x} \in \operatorname{Ker} A^+$, eli $A^+\boldsymbol{x} = \Theta$, niin kertomalla vasemmalta matriisilla A saamme $AA^+\boldsymbol{x} = \Theta$, joten Ker $A^+ \subseteq$ Ker (AA^+). Jos taas $\boldsymbol{x} \in$ Ker (AA^+), eli $AA^+\boldsymbol{x} = \Theta$, niin kertomalla tämä yhtälö vasemmalta matriisilla A^+ saadaan

$$A^+AA^+\boldsymbol{x} = A^+\boldsymbol{x} = \Theta,$$

joten Ker $AA^+ \subseteq$ Ker A^+. Täten siis Ker $(AA^+) =$ Ker A^+.

Osoitetaan sitten, että Ran $(AA^+) =$ Ran A. Jos $\boldsymbol{x} \in$ Ran A, eli $A\boldsymbol{y} = \boldsymbol{x}$, jollakin $\boldsymbol{y} \in \mathbb{C}^n$, niin määritelmän 6.1 ja lauseen 6.5 nojalla saamme

$$(AA^+)A\boldsymbol{y} = A\boldsymbol{y} = \boldsymbol{x},$$

joten Ran $A \subseteq$ Ran (AA^+). Jos taasen $\boldsymbol{x} \in$ Ran (AA^+), eli $AA^+\boldsymbol{y} = \boldsymbol{x}$, jollakin $\boldsymbol{y} \in \mathbb{C}^n$, niin selvästi $\boldsymbol{x} \in$ Ran A, joten Ran $(AA^+) \subseteq$ Ran A. Nyt siis Ran $(AA^+) =$ Ran A. □

Korvaamalla matriisin A matriisilla A^+ saamme myös tulokset

$$\text{Ker } (A^+A) = \text{Ker } A \qquad \text{ja} \qquad \text{Ran } (A^+A) = \text{Ran } A^+.$$

Lause 6.8. *Olkoon $A \in \mathbb{C}^{m\times n}$ ja olkoon $A^+ \in \mathbb{C}^{n\times m}$ matriisin A pseudoinverssi. Tällöin*

$$\text{Ker}\, A^+ = (\text{Ran}\, A)^\perp \qquad \textit{ja} \qquad \text{Ran}\, A^+ = (\text{Ker}\, A)^\perp$$

Todistus. Osoitetaan ensin, että Ran $A^+ =$ (Ker $A)^\perp$. Oletetaan, että $\boldsymbol{x} \in$ Ker A ja $\boldsymbol{y} \in$ Ran A^+, eli $A\boldsymbol{x} = \Theta$ ja $A^+\boldsymbol{z} = \boldsymbol{y}$, jollakin $\boldsymbol{z} \in \mathbb{C}^n$. Meidän tulee siis osoittaa, että $\langle \boldsymbol{y}, \boldsymbol{x}\rangle = 0$. Nyt määritelmän 6.1 avulla saamme

$$\begin{aligned}\langle \boldsymbol{x}, \boldsymbol{y}\rangle &= \boldsymbol{y}^*\boldsymbol{x} = (A^+\boldsymbol{z})^*\boldsymbol{x} = (A^+AA^+\boldsymbol{z})^*\boldsymbol{x} = \boldsymbol{z}^*(A^+)^*A^*(A^+)^*\boldsymbol{x}\\ &= \boldsymbol{z}^*(A^+)^*(A^+A)^*\boldsymbol{x} = \boldsymbol{z}^*(A^+)^*A^+A\boldsymbol{x} = \boldsymbol{z}^*(A^+)^*A^+\Theta = 0,\end{aligned}$$

joten $\boldsymbol{y} \in$ (Ker $A)^\perp$, eli Ran $A^+ \subseteq$ (Ker $A)^\perp$.

Oletetaan sitten, että $\boldsymbol{y} \in$ (Ker $A)^\perp$. Tällöin $\boldsymbol{y} \in$ Ran A^*, koska lauseen 2.15 nojalla (Ker $A)^\perp =$ Ran A^*. Tällöin lauseen 6.5 nojalla $A^+A\boldsymbol{y} = \boldsymbol{y}$, joten $\boldsymbol{y} \in$ Ran A^+. Nyt (Ker $A)^\perp \subseteq$ Ran A^+, joten olemme todistaneet, että Ran $A^+ =$ (Ker $A)^\perp$.

Osoitetaan seuraavaksi, että Ker $A^+ =$ (Ran $A)^\perp$. Oletetaan, että $\boldsymbol{x} \in$ Ker A^+ ja $\boldsymbol{y} \in$ Ran A, jolloin $A^+\boldsymbol{x} = \Theta$ ja $A\boldsymbol{z} = \boldsymbol{y}$ jollakin $\boldsymbol{z} \in \mathbb{C}^n$. Nyt

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{y}^*\boldsymbol{x} = (AA^+A\boldsymbol{z})^*\boldsymbol{x} = \boldsymbol{z}^*A^*(AA^+)^*\boldsymbol{x} = \boldsymbol{z}^*A^*AA^+\boldsymbol{x} = 0,$$

joten $\boldsymbol{x} \in$ (Ran $A)^\perp$, eli Ker $A^+ \subseteq$ (Ran $A)^\perp$. Oletetaan sitten, että $\boldsymbol{x} \in$ (Ran $A)^\perp$, jolloin lauseen 2.15 nojalla $\boldsymbol{x} \in$ Ker A^*. Tällöin siis $A^*\boldsymbol{x} = \Theta$, ja jos kerromme tämän yhtälön molemmat puolet vasemmalta matriisilla $(A^+)^*$, saamme $(A^+)^*A^*\boldsymbol{x} = (AA^+)^*\boldsymbol{x} = AA^+\boldsymbol{x} = \Theta$, jolloin siis $\boldsymbol{x} \in$ Ker AA^+. Koska lauseen 6.7 nojalla Ker $AA^+ =$ Ker A^+, niin nyt $\boldsymbol{x} \in$ Ker A^+, joten (Ran $A)^\perp \subseteq$ Ker A^+. Olemme siis todistaneet, että Ker $A^+ =$ (Ran $A)^\perp$. □

Nyt lauseiden 2.15 ja 6.8 avulla saamme mielenkiintoisen tuloksen. Yhdistämällä lauseiden tulokset huomaamme, että

$$\text{Ker } A^* = \text{Ker } A^+ = (\text{Ran } A)^\perp$$

ja

$$\text{Ran } A^* = \text{Ran } A^+ = (\text{Ker } A)^\perp.$$

6.2 Pienimmän neliösumman probleema

Tässä kappaleessa käsittelemme pseudoinverssin soveltamista lineaarisen systeemin $A\boldsymbol{x} = \boldsymbol{b}$ ratkaisemiseen. Joitakin aihetta käsitteleviä esitietoja löytyy aliluvusta 2.5.

Kyseisessä aliluvussa esitimme, että vektori $\boldsymbol{x} \in \mathbb{C}^n$ on lineaarisen systeemin ratkaisu, jos ja vain jos $\|A\boldsymbol{x} - \boldsymbol{b}\| = 0$. Toistaalta esitimme myös, että tällainen vektori on mahdollista löytää, jos ja vain jos $\boldsymbol{b} \in \text{Ran } A$. Jos siis $\boldsymbol{b} \notin \text{Ran } A$, niin lineaarisella systeemillä ei ole ratkaisuja, jolloin myös $\|A\boldsymbol{x} - \boldsymbol{b}\| > 0$. Vaikka ratkaisuja ei löydy, voimme kuitenkin etsiä vektoria, jolla lauseke $\|A\boldsymbol{x} - \boldsymbol{b}\|$ saa mahdollisimman pienen arvon. Tätä kutsutaan lineaarisen yhtälöryhmän ratkaisun approksimoinniksi, ja tästä aiheesta lisää seuraavassa.

Määritelmä 6.2. Olkoon $A\boldsymbol{x} = \boldsymbol{b}$ lineaarinen systeemi, jossa $A \in \mathbb{C}^{m\times n}$, $\boldsymbol{x} \in \mathbb{C}^n$ ja $\boldsymbol{b} \in \mathbb{C}^m$. Tällöin vektoria $\boldsymbol{x_0}$, jolle

$$\|A\boldsymbol{x_0} - \boldsymbol{b}\| = \min \|A\boldsymbol{x} - \boldsymbol{b}\|, \tag{6.1}$$

kun $\boldsymbol{x} \in \mathbb{C}^n$, kutsutaan *lineaarisen systeemin $A\boldsymbol{x} = \boldsymbol{b}$ pienimmän neliösumman ratkaisuksi.* Jos vielä $\|\boldsymbol{x_0}\| \leq \|\boldsymbol{y}\|$ kaikilla vektoreilla $\boldsymbol{y} \in \mathbb{C}^n$, jotka ovat pienimmän neliösumman ratkaisuja, niin kutsutaan vektoria $\boldsymbol{x_o}$ *lineaarisen systeemin $A\boldsymbol{x} = \boldsymbol{b}$ parhaaksi pienimmän neliösumman ratkaisuksi.*

Tällöin vektorilla $\boldsymbol{x_0}$ on lyhin mahdollinen euklidinen normi niistä vektoreista, jotka toteuttavat yhtälön (6.1). Osoitamme seuraavaksi, että vektori $\boldsymbol{x_0} = A^+\boldsymbol{b}$ on edelliset ehdot täyttävä vektori eli paras pienimmän neliösumman ratkaisu. Jos siis $\boldsymbol{b} \in \text{Ran } A$, niin $\boldsymbol{x_0} = A^+\boldsymbol{b}$ on lyhimmän normin omaava vektori ratkaisuvektorien joukosta tai yksikäsitteinen ratkaisu. Jos taas $\boldsymbol{b} \notin \text{Ran } A$, niin $\boldsymbol{x_0} = A^+\boldsymbol{b}$ on se yksikäsitteinen vektori, joka minimoi lausekkeen $\|A\boldsymbol{x} - \boldsymbol{b}\|$.

Lause 6.9. *Lineearisen systeemin $A\boldsymbol{x} = \boldsymbol{b}$, jossa $A \in \mathbb{C}^{m\times n}$, $\boldsymbol{x} \in \mathbb{C}^n$ ja $\boldsymbol{b} \in \mathbb{C}^m$, paras pienimmän neliösumman ratkaisu on vektori $\boldsymbol{x_0} = A^+\boldsymbol{b}$.*

Todistus. Mille tahansa vektorille $\boldsymbol{x} \in \mathbb{C}^n$ saamme

$$A\boldsymbol{x} - \boldsymbol{b} = A\boldsymbol{x} - AA^+\boldsymbol{b} + AA^+\boldsymbol{b} - \boldsymbol{b} = A(\boldsymbol{x} - A^+\boldsymbol{b}) + (AA^+ - I)\boldsymbol{b}.$$

Selvästi $A(\boldsymbol{x} - A^+\boldsymbol{b}) \in \text{Ran } A$. Osoitetaan seuraavaksi, että $(AA^+ - I)\boldsymbol{b} \in (\text{Ran } A)^\perp$, eli $\langle (AA^+ - I)\boldsymbol{b}, \boldsymbol{y}\rangle = 0$ kaikille vektoreille $\boldsymbol{y} \in \text{Ran } A$. Nyt määritelmän 6.1 ja lauseen 6.5 avulla saamme

$$\begin{aligned}\Big\langle (AA^+ - I)\boldsymbol{b}, \boldsymbol{y}\Big\rangle &= \boldsymbol{y}^*(AA^+ - I)\boldsymbol{b} = \boldsymbol{y}^*AA^+\boldsymbol{b} - \boldsymbol{y}^*\boldsymbol{b} \\ &= ((AA^+)^*\boldsymbol{y})^*\boldsymbol{b} - \boldsymbol{y}^*\boldsymbol{b} = (AA^+\boldsymbol{y})^*\boldsymbol{b} - \boldsymbol{y}^*\boldsymbol{b} \\ &= \boldsymbol{y}^*\boldsymbol{b} - \boldsymbol{y}^*\boldsymbol{b} = 0,\end{aligned}$$

joten $(AA^+ - I)\boldsymbol{b} \in (\text{Ran } A)^\perp$. Koska nyt vektorit $A(\boldsymbol{x} + A^+\boldsymbol{b})$ ja $(AA^+ - I)\boldsymbol{b}$ ovat ortogonaalisia, saamme Pythagoraan lauseen (lause 2.10) nojalla

$$\begin{aligned}\|A\boldsymbol{x} - \boldsymbol{b}\|^2 &= \Big\|A(\boldsymbol{x} - A^+\boldsymbol{b})\Big\|^2 + \Big\|(AA^+ - I)\boldsymbol{b}\Big\|^2 \\ &= \|A(\boldsymbol{x} - \boldsymbol{x_0})\|^2 + \Big\|AA^+\boldsymbol{b} - \boldsymbol{b}\Big\|^2 = \|A(\boldsymbol{x} - \boldsymbol{x_0})\|^2 + \|A\boldsymbol{x_0} - \boldsymbol{b}\|^2 .\end{aligned}$$

Nyt siis $\boldsymbol{x_0}$ on pienimmän neliösumman ratkaisu lineaariselle systeemille $A\boldsymbol{x} = \boldsymbol{b}$.

Osoitetaan vielä, että $\|\boldsymbol{x_0}\| \leq \|\boldsymbol{x_1}\|$ kaikilla vektoreilla $\boldsymbol{x_1} \in \mathbb{C}^n$, jotka ovat pienimmän neliösumman ratkaisuja lineaariselle systeemille $A\boldsymbol{x} = \boldsymbol{b}$. Jos sijoitamme edelliseen yhtälöön $\boldsymbol{x} = \boldsymbol{x_1}$, saamme

$$\|A(\boldsymbol{x_1} - \boldsymbol{x_0})\|^2 + \|A\boldsymbol{x_0} - \boldsymbol{b}\|^2 = \|A\boldsymbol{x_1} - \boldsymbol{b}\|^2 ,$$

josta seuraa, että

$$\|A(\boldsymbol{x_1} - \boldsymbol{x_0})\|^2 = 0,$$

koska $\|A\boldsymbol{x_0} - \boldsymbol{b}\|^2 = \|A\boldsymbol{x_1} - \boldsymbol{b}\|^2$. Tällöin siis tulee olla $A(\boldsymbol{x_1} - \boldsymbol{x_0}) = \Theta$, joten $(\boldsymbol{x_1} - \boldsymbol{x_0}) \in \text{Ker } A$. Nyt $\boldsymbol{x_0} = A^+\boldsymbol{b} \in \text{Ran } A^+$ ja lauseen 6.8 nojalla Ker A ja Ran A^+ ovat toistensa ortogonaalikomplementteja, joten $\boldsymbol{x_0}$ ja $(\boldsymbol{x_1} - \boldsymbol{x_0})$ ovat ortogonaalisia. Voimme taas käyttää Pythagoraan lausetta, jolla saamme

$$\|\boldsymbol{x_1}\|^2 = \|\boldsymbol{x_0} + (\boldsymbol{x_1} - \boldsymbol{x_0})\|^2 = \|\boldsymbol{x_0}\|^2 + \|\boldsymbol{x_1} - \boldsymbol{x_0}\|^2 \geq \|\boldsymbol{x_0}\|^2 ,$$

joten olemme todistaneet lauseen. □

Koska matriisi A^+ on aina yksikäsitteisesti määritelty, niin myös paras pienimmän neliösumman ratkaisu on myös aina yksikäsitteisesti määritelty. Täten siis kaikki muut parhaat pienimmän neliösumman ratkaisut palautuvat pseudoinverssin käsitteeseen. Jos siis $A \in \mathbb{C}^{n\times n}$ on kääntyvä, niin $A^+ = A^{-1}$ ja $\boldsymbol{x_0} = A^{-1}\boldsymbol{b}$. Myös lauseessa 6.3 esitetyistä muodoista $A^+ = A^*(A^*A)^{-1}$ ja $A^+ = (AA^*)^{-1}A^*$ saadaan parhaat pienimmän neliösumman ratkaisut $\boldsymbol{x_0} = A^*(A^*A)^{-1}\boldsymbol{b}$ ja $\boldsymbol{x_0} = (AA^*)^{-1}A^*\boldsymbol{b}$ lauseessa esitettyjen ehtojen mukaisesti.

Esimerkki 6.2. Tarkastellaan lineaarista systeemiä $A\boldsymbol{x} = \boldsymbol{b}$, jossa

$$A = \begin{bmatrix} -3 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \qquad \text{ja} \qquad \boldsymbol{b} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$

Lineaarisen systeemin paras pienimmän neliösumman ratkaisu saadaan nyt lauseen 6.9 nojalla kaavasta $\boldsymbol{x_o} = A^+\boldsymbol{b}$, joten muodostetaan matriisi A^+ matriisin A singulaariarvohajotelman avulla. Nyt

$$A^*A = \begin{bmatrix} -3 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 0 \\ 0 & 6 \end{bmatrix}.$$

Tällöin siis matriisin A singulaariarvot ovat $s_1 = \sqrt{11}$ ja $s_2 = \sqrt{6}$. Voimme valita matriisin A^*A ortonormaaliksi kannaksi vektoriavaruuden $\mathbb{C}^2$ luonnollisen kannan, jolloin

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Määrittääksemme matriisin U meidän tulee selvittää matriisin AA^* ominaisavaruuden ortonormaali kanta. Nyt lauseen 5.4 nojalla

$$\boldsymbol{y_1} = \frac{A\boldsymbol{x_1}}{s_1} = \frac{\begin{bmatrix} -3 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{\sqrt{11}} = \frac{1}{\sqrt{11}} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

ja

$$\boldsymbol{y_2} = \frac{A\boldsymbol{x_2}}{s_2} = \frac{\begin{bmatrix} -3 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}}{\sqrt{6}} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

Koska vektoreiden $\boldsymbol{y_1}, \boldsymbol{y_2}, \boldsymbol{y_3}$ tulee muodostaa vektoriavaruuden $\mathbb{C}^3$ ortonormaali kanta, voimme helpohkoilla laskutoimituksilla ratkaista myös vektorin $\boldsymbol{y_3}$, josta saamme

$$\boldsymbol{y_3} = \frac{1}{\sqrt{66}} \begin{bmatrix} 1 \\ -4 \\ 7 \end{bmatrix}.$$

Tällöin siis matriisi

$$U = \begin{bmatrix} \frac{-3}{\sqrt{11}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{66}} \\ \frac{1}{\sqrt{11}} & \frac{2}{\sqrt{6}} & \frac{-4}{\sqrt{66}} \\ \frac{1}{\sqrt{11}} & \frac{1}{\sqrt{6}} & \frac{7}{\sqrt{66}} \end{bmatrix}.$$

Matriisin A singulaariarvohajotelma on

$$A = \begin{bmatrix} -3 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{-3}{\sqrt{11}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{66}} \\ \frac{1}{\sqrt{11}} & \frac{2}{\sqrt{6}} & \frac{-4}{\sqrt{66}} \\ \frac{1}{\sqrt{11}} & \frac{1}{\sqrt{6}} & \frac{7}{\sqrt{66}} \end{bmatrix} \begin{bmatrix} \sqrt{11} & 0 \\ 0 & \sqrt{6} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = UDV^*,$$

joten matriisin A pseudoinverssi on

$$A^+ = VD^+U^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{11}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} \frac{-3}{\sqrt{11}} & \frac{1}{\sqrt{11}} & \frac{1}{\sqrt{11}} \\ \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{66}} & \frac{-4}{\sqrt{66}} & \frac{7}{\sqrt{66}} \end{bmatrix} = \begin{bmatrix} -\frac{3}{11} & \frac{1}{11} & \frac{1}{11} \\ \frac{1}{6} & \frac{2}{6} & \frac{1}{6} \end{bmatrix}.$$

Tällöin lineaarisen systeemin $A\boldsymbol{x} = \boldsymbol{b}$ paras pienimmän neliösumman ratkaisu on

$$\boldsymbol{x_o} = A^+\boldsymbol{b} = \begin{bmatrix} -\frac{3}{11} & \frac{1}{11} & \frac{1}{11} \\ \frac{1}{6} & \frac{2}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{6}{11} - \frac{1}{11} + 0 \\ \frac{2}{6} - \frac{2}{6} + 0 \end{bmatrix} = \begin{bmatrix} -\frac{7}{11} \\ 0 \end{bmatrix}.$$

Koska rank $A = 2$, jolloin lauseen 6.3 ehdot täyttyvät, niin olisimme voineet laskea parhaan pienimmän neliösumman ratkaisun myös kaavalla

$$\boldsymbol{x_o} = A^+\boldsymbol{b} = (A^*A)^{-1}A^*\boldsymbol{b} = \begin{bmatrix} \frac{1}{11} & 0 \\ 0 & \frac{1}{6} \end{bmatrix} \begin{bmatrix} -3 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{7}{11} \\ 0 \end{bmatrix}.$$

Jos tiedämme matriisin singulaariarvot ja oikeanpuoleisen singulaarikannan (eli matriisin A^*A ortonormaalin ominaisvektorikannan), voimme muodostaa myös toisenlaisen muodon parhaan pienimmän neliösumman ratkaisulle.

Lause 6.10. *Olkoon $A\boldsymbol{x} = \boldsymbol{b}$ lineaarisella systeemi, jossa $A \in \mathbb{C}^{m\times n}$, $\boldsymbol{b} \in \mathbb{C}^m$ ja $\boldsymbol{x} \in \mathbb{C}^n$. Muodostakoot vektorit $\boldsymbol{x_1}, \dots, \boldsymbol{x_n}$ matriisin A oikeanpuoleisen singulaarikannan ja olkoot matriisin A singulaariarvot $s_1, \dots, s_r$. Jos merkitään*

$$\alpha_i = \frac{\langle A^*\boldsymbol{b}, \boldsymbol{x_i} \rangle}{s_i^2},$$

kun $i = 1, 2, \dots, r$, niin lineaarisen systeemin paras pienimmän neliösumman ratkaisu on

$$\boldsymbol{x_o} = \alpha_1\boldsymbol{x_1} + \alpha_2\boldsymbol{x_2} + \dots + \alpha_r\boldsymbol{x_r}.$$

Todistus. Muodostakoot vektorit $\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_m}$ matriisin A vasemmanpuoleisen singulaarikannan. Nyt vektori $\boldsymbol{b}$ voidaan esittää lauseen 2.14 nojalla näiden singulaarivektoreiden lineaarikombinaationa, eli

$$\boldsymbol{b} = \beta_1\boldsymbol{y_1} + \beta_2\boldsymbol{y_2} + \dots \beta_m\boldsymbol{y_m},$$

jossa $\beta_i = \langle \boldsymbol{b}, \boldsymbol{y_i} \rangle$, kun $i = 1, 2, \dots, m$. Nyt voimme kirjottaa lauseen 6.6 nojalla

$$\boldsymbol{x_0} = A^+\boldsymbol{b} = \sum_{i=1}^{m} \beta_i A^+ \boldsymbol{y_i} = \sum_{i=1}^{r} \frac{\beta_i}{s_i} \boldsymbol{x_i}.$$

Toisaalta lauseen 5.4 nojalla $\boldsymbol{y_i} = s_i^{-1} A\boldsymbol{x_i}$, joten

$$\frac{\beta_i}{s_i} = \frac{\langle \boldsymbol{b}, \boldsymbol{y_i} \rangle}{s_i} = \frac{\boldsymbol{y_i^*b}}{s_i} = \frac{(A\boldsymbol{x_i})^*\boldsymbol{b}}{s_i^2} = \frac{\langle \boldsymbol{b}, A\boldsymbol{x_i} \rangle}{s_i^2} = \frac{\langle A^*\boldsymbol{b}, \boldsymbol{x_i} \rangle}{s_i^2}.$$

Täten saamme

$$\boldsymbol{x_0} = \sum_{i=1}^{r} \frac{\beta_i}{s_i} \boldsymbol{x_i} = \sum_{i=1}^{r} \frac{\langle A^*\boldsymbol{b}, \boldsymbol{x_i} \rangle}{s_i^2} \boldsymbol{x_i},$$

joka on haluttu muoto. □

Esimerkki 6.3. Ratkaistaan nyt esimerkin 6.2 probleema käyttämällä lausetta 6.10. Esimerkissä 6.2 ratkaisimme vektorit $\boldsymbol{x_1}$ ja $\boldsymbol{x_2}$, jotka muodostivat vektoriavaruuden $\mathbb{C}^2$ luonnollisen kannan. Koska

$$A^*\boldsymbol{b} = \begin{bmatrix} -3 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -7 \\ 0 \end{bmatrix},$$

niin saamme parhaaksi pienemmän neliösumman ratkaisuksi

$$\boldsymbol{x_o} = \frac{\boldsymbol{x_1^*}A^*\boldsymbol{b}}{s_1^2}\boldsymbol{x_1} + \frac{\boldsymbol{x_2^*}A^*\boldsymbol{b}}{s_2^2}\boldsymbol{x_2} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} -7 \\ 0 \end{bmatrix}}{11} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} -7 \\ 0 \end{bmatrix}}{6} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{7}{11} \\ 0 \end{bmatrix},$$

joka on sama vastaus kuin saimme esimerkissä 6.2.

Viitteet

[1] R.A. Horn & C.R. Johnson, *Matrix Analysis*, 1990, Cambridge University Press, ISBN 0-521-38632-2.

[2] S.K. Kivelä, *Matriisilasku ja Lineaarialgebra*, 1984, Otatieto, 7.painos ISBN 951-671-368-8.

[3] P. Lancaster & M. Tismenetsky, *The Theory of Matrices: Second edition with applications*, 1985, Academic press, Inc., Orlando Florida, ISBN 0-12-435560-9.

[4] P.J. Olver & C. Shakiban, *Applied Linear Algebra*, 2006, Pearson Education, 1.painos ISBN 0-13-147382-4.